CALCULUS ILLUSTRATED

VOLUME 2: DIFFERENTIAL CALCULUS

PETER SAVELIEV

To the student

Mathematics is a science. Just as the rest of the scientists, mathematicians are trying to understand how the Universe operates and discover its laws. When successful, they write these laws as short statements called "theorems". In order to present these laws conclusively and precisely, a dictionary of the new concepts is also developed; its entries are called "definitions". These two make up the most important part of any mathematics book.

This is how definitions, theorems, and some other items are used as building blocks of the scientific theory we present in this text.

Every new concept is introduced with utmost specificity.

Definition 0.0.1: square root

Suppose a is a positive number. Then the *square root* of a is a positive number x, such that $x^2 = a$.

The term being introduced is given in *italics*. The definitions are then constantly referred to throughout the text.

New symbolism may also be introduced.

Square root

$$\sqrt{a}$$

Consequently, the notation is freely used throughout the text.

We may consider a specific instance of a new concept either before or after it is explicitly defined.

Example 0.0.2: length of diagonal

What is the length of the diagonal of a 1×1 square? The square is made of two right triangles and the diagonal is their shared hypotenuse. Let's call it a. Then, by the *Pythagorean Theorem*, the square of a is $1^2 + 1^2 = 2$. Consequently, we have:

$$a^2 = 2\,.$$

We immediately see the need for the square root! The length is, therefore, $a = \sqrt{2}$.

You can skip some of the examples without violating the flow of ideas, at your own risk.

All new material is followed by a few little tasks, or questions, like this.

Exercise 0.0.3

Find the height of an equilateral triangle the length of the side of which is 1.

The exercises are to be attempted (or at least considered) immediately.

Most of the in-text exercises are not elaborate. They aren't, however, entirely routine as they require understanding of, at least, the concepts that have just been introduced. Additional exercise *sets* are placed in the appendix as well as at the book's website: calculus123.com. Do not start your study with the exercises! Keep in mind that the exercises are meant to test – indirectly and imperfectly – how well the *concepts* have been learned.

There are sometimes words of caution about common mistakes made by the students.

Warning!

In spite of the fact that $(-1)^2 = 1$, there is only one square root of 1, $\sqrt{1} = 1$.

The most important facts about the new concepts are put forward in the following manner.

Theorem 0.0.4: Product of Roots

For any two positive numbers a and b, we have the following identity:

$$\boxed{\sqrt{a} \cdot \sqrt{b} = \sqrt{a \cdot b}}$$

The theorems are constantly referred to throughout the text.

As you can see, theorems may contain formulas; a theorem supplies limitations on the applicability of the formula it contains. Furthermore, every formula is a part of a theorem, and using the former without knowing the latter is perilous.

There is no need to memorize definitions or theorems (and formulas), initially. With enough time spent with the material, the main ones will eventually become familiar as they continue to reappear in the text. Watch for words "important", "crucial", etc. Those new concepts that do not reappear in this text are likely to be seen in the next mathematics book that you read. You need to, however, be aware of all of the definitions and theorems and be able to find the right one when necessary.

Often, but not always, a theorem is followed by a thorough argument as a justification.

Proof.

Suppose $A = \sqrt{a}$ and $B = \sqrt{b}$. Then, according to the definition, we have the following:

$$a = A^2 \quad \text{and} \quad b = B^2 \ .$$

Therefore, we have:

$$a \cdot b = A^2 \cdot B^2 = A \cdot A \cdot B \cdot B = (A \cdot B) \cdot (A \cdot B) = (AB)^2 \ .$$

Hence, $\sqrt{ab} = A \cdot B$, again according to the definition.

Some proofs can be skipped at first reading.

Its highly detailed exposition makes the book a good choice for *self-study*. If this is your case, these are my suggestions.

While reading the book, try to make sure that you understand new concepts and ideas. Keep in mind, however, that some are more important that others; they are marked accordingly. Come back (or jump forward) as needed. Contemplate. Find other sources if necessary. You should not turn to the exercise sets until you have become comfortable with the material.

What to do about exercises when solutions aren't provided? First, use the examples. Many of them contain a problem – with a solution. Try to solve the problem – before or after reading the solution. You can also find exercises online or make up your own problems and solve them!

I strongly suggest that your solution should be thoroughly *written*. You should write in complete sentences, including all the algebra. For example, you should appreciate the difference between these two:

Wrong: $\boxed{\begin{array}{l} 1+1 \\ 2 \end{array}}$ Right: $\boxed{\begin{array}{l} 1+1 \\ =2 \end{array}}$

The latter reads "one added to one is two", while the former cannot be read. You should also justify all your steps and conclusions, including all the algebra. For example, you should appreciate the difference between these two:

Wrong: $\boxed{\begin{array}{l} 2x = 4 \\ x = 2 \end{array}}$ Right: $\boxed{\begin{array}{l} 2x = 4; \text{ therefore,} \\ x = 2\,. \end{array}}$

The standards of thoroughness are provided by the examples in the book.

Next, your solution should be thoroughly *read*. This is the time for self-criticism: Look for errors and weak spots. It should be re-read and then rewritten. Once you are convinced that the solution is correct and the presentation is solid, you may show it to a knowledgeable person for a once-over.

Next, you may turn to modeling projects. Spreadsheets (Microsoft Excel or similar) are chosen to be used for graphing and modeling. One can achieve as good results with packages specifically designed for these purposes, but spreadsheets provide a tool with a wider scope of applications. Programming is another option.

Good luck!

To the teacher

The bulk of the material in the book comes from my lecture notes.

There is little emphasis on closed-form computations and algebraic manipulations. I do think that a person who has never integrated by hand (or differentiated, or applied the quadratic formula, etc.) cannot possibly understand integration (or differentiation, or quadratic functions, etc.). However, a large proportion of time and effort can and should be directed toward:

- understanding of the concepts and
- modeling in realistic settings.

The challenge of this approach is that it requires more abstraction rather than less.

Visualization is the main tool used to deal with this challenge. Illustrations are provided for every concept, big or small. The pictures that come out are sometimes very precise but sometimes serve as mere metaphors for the concepts they illustrate. The hope is that they will serve as visual "anchors" in addition to the words and formulas.

It is unlikely that a person who has never plotted the graph of a function by hand can understand graphs or functions. However, what if we want to plot more than just a few points in order to visualize curves, surfaces, vector fields, etc.? Spreadsheets were chosen over graphic calculators for visualization purposes because they represent the shortest step away from pen and paper. Indeed, the data is plotted in the simplest manner possible: one cell - one number - one point on the graph. For more advanced tasks such as modeling, spreadsheets were chosen over other software and programming options for their wide availability and, above all, their simplicity. Nine out of ten, the spreadsheet shown was initially created from scratch in front of the students who were later able to follow my footsteps and create their own.

About the tests. The book isn't designed to prepare the student for some preexisting exam; on the contrary, assignments should be based on what has been learned. The students' understanding of the concepts needs to be tested but, most of the time, this can be done only indirectly. Therefore, a certain share of routine, mechanical problems is inevitable. Nonetheless, no topic deserves more attention just because it's likely to be on the test.

If at all possible, don't make the students memorize formulas.

In the order of topics, the main difference from a typical calculus textbook is that sequences come before everything else. The reasons are the following:

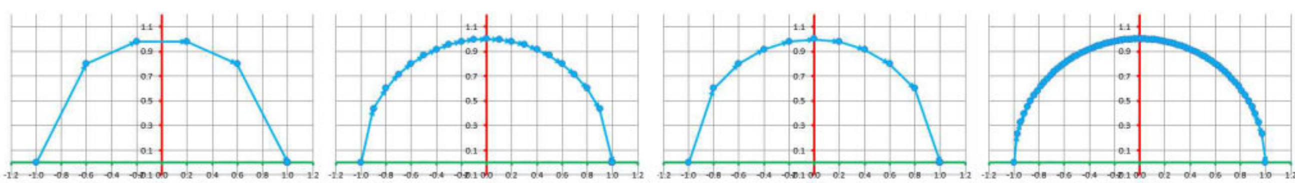
- Sequences are the simplest kind of functions.
- Limits of sequences are simpler than limits of general functions (including the ones at infinity).
- The sigma notation, the Riemann sums, and the Riemann integral make more sense to a student with a solid background in sequences.
- A quick transition from sequences to series often leads to confusion between the two.
- Sequences are needed for modeling, which should start as early as possible.

From the discrete to the continuous

It's no secret that a vast majority of calculus students will never use what they have learned. Poor career choices aside, a former calculus student is often unable to recognize the mathematics that is supposed to surround him. Why does this happen?

Calculus is the science of change. From the very beginning, its peculiar challenge has been to study and measure *continuous* change: curves and motion along curves. These curves and this motion are represented by *formulas*. Skillful manipulation of those formulas is what solves calculus problems. For over 300 years, this approach has been extremely successful in sciences and engineering. The successes are well-known: projectile motion, planetary motion, flow of liquids, heat transfer, wave propagation, etc. Teaching calculus follows this approach: An overwhelming majority of what the student does is manipulation of formulas on a piece of paper. But this means that all the problems the student faces were (or could have been) solved in the 18th or 19th centuries!

This isn't good enough anymore. What has changed since then? The computers have appeared, of course, and computers don't manipulate formulas. They don't help with solving – in the traditional sense of the word – those problems from the past centuries. Instead of *continuous*, computers excel at handling *incremental* processes, and instead of formulas they are great at managing discrete (digital) data. To utilize these advantages, scientists "discretize" the results of calculus and create algorithms that manipulate the digital data. The solutions are approximate but the applicability is unlimited. Since the 20th century, this approach has been extremely successful in sciences and engineering: aerodynamics (airplane and car design), sound and image processing, space exploration, structure of the atom and the universe, etc. The approach is also circuitous: Every concept in calculus *starts* – often implicitly – as a discrete approximation of a continuous phenomenon!



Calculus is the science of change, *both* incremental and continuous. The former part – the so-called discrete calculus – may be seen as the study of incremental phenomena and the quantities *indivisible* by their very nature: people, animals, and other organisms, moments of time, locations of space, particles, some commodities, digital images and other man-made data, etc. With the help of the calculus machinery called "limits", we invariably choose to transition to the continuous part of calculus, especially when we face continuous phenomena and the quantities *infinitely divisible* either by their nature or by assumption: time, space, mass, temperature, money, some commodities, etc. Calculus produces definitive results and absolute accuracy – but only for problems amenable to its methods! In the classroom, the problems are simplified until they become manageable; otherwise, we circle back to the discrete methods in search of approximations.

Within a typical calculus course, the student simply never gets to complete the "circle"! Later on, the graduate is likely to think of calculus only when he sees formulas and rarely when he sees numerical data.

In this book, every concept of calculus is first introduced in its discrete, "pre-limit", incarnation – elsewhere typically hidden inside proofs – and then used for modeling and applications well before its continuous counterpart emerges. The properties of the former are discovered first and then the matching properties of the latter are found by making the increment smaller and smaller, at the *limit*:

$$\boxed{\begin{array}{c}\text{discrete}\\ \text{calculus}\end{array}} \xrightarrow{\ \ \Delta x\to 0\ \ } \boxed{\begin{array}{c}\text{continuous}\\ \text{calculus}\end{array}}$$

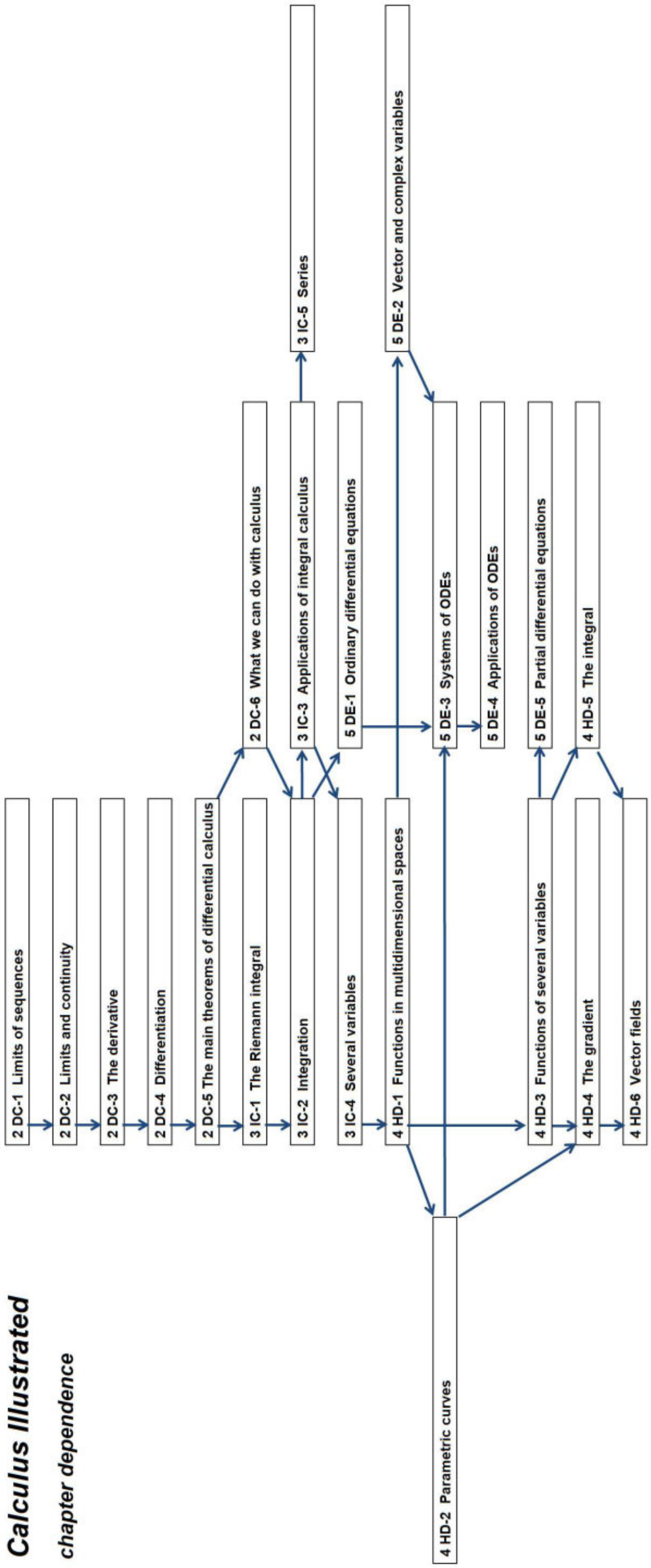
The volume and chapter references for *Calculus Illustrated*

This book is a part of the series *Calculus Illustrated*. The series covers the standard material of the undergraduate calculus with a substantial review of precalculus and a preview of elementary ordinary and partial differential equations. Below is the list of the books of the series, their chapters, and the way the present book (parenthetically) references them.

	■ **Calculus Illustrated. Volume 1: Precalculus**
1 PC-1	Calculus of sequences
1 PC-2	Sets and functions
1 PC-3	Compositions of functions
1 PC-4	Classes of functions
1 PC-5	Algebra and geometry
	■ **Calculus Illustrated. Volume 2: Differential Calculus**
2 DC-1	Limits of sequences
2 DC-2	Limits and continuity
2 DC-3	The derivative
2 DC-4	Differentiation
2 DC-5	The main theorems of differential calculus
2 DC-6	What we can do with calculus
	■ **Calculus Illustrated. Volume 3: Integral Calculus**
3 IC-1	The Riemann integral
3 IC-2	Integration
3 IC-3	Applications of integral calculus
3 IC-4	Several variables
3 IC-5	Series
	■ **Calculus Illustrated. Volume 4: Calculus in Higher Dimensions**
4 HD-1	Functions in multidimensional spaces
4 HD-2	Parametric curves
4 HD-3	Functions of several variables
4 HD-4	The gradient
4 HD-5	The integral
4 HD-6	Vector fields
	■ **Calculus Illustrated. Volume 5: Differential Equations**
5 DE-1	Ordinary differential equations
5 DE-2	Vector and complex variables
5 DE-3	Systems of ODEs
5 DE-4	Applications of ODEs
5 DE-5	Partial differential equations

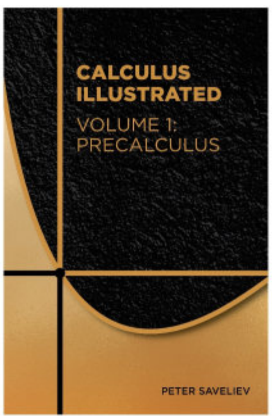
The volumes can be read independently.

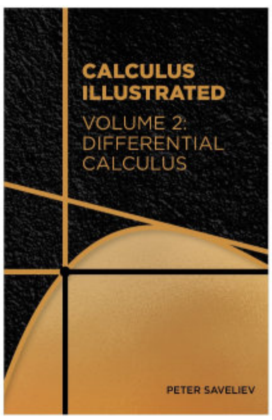
A possible sequence of chapters is presented below. An arrow from A to B means that chapter B shouldn't be read before chapter A.

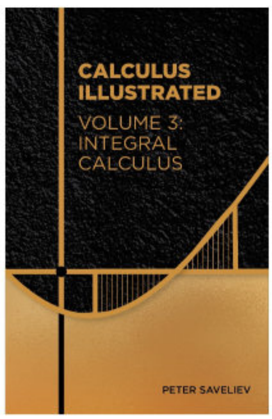


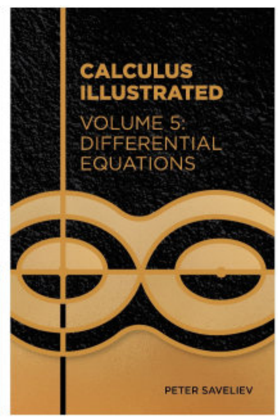
About the author

Peter Saveliev is a professor of mathematics at Marshall University, Huntington, West Virginia, USA. After a Ph.D. from the University of Illinois at Urbana-Champaign, he devoted the next 20 years to teaching mathematics. Peter is the author of a graduate textbook *Topology Illustrated* published in 2016. He has also been involved in research in algebraic topology and several other fields. His non-academic projects have been: digital image analysis, automated fingerprint identification, and image matching for missile navigation/guidance.









Contents

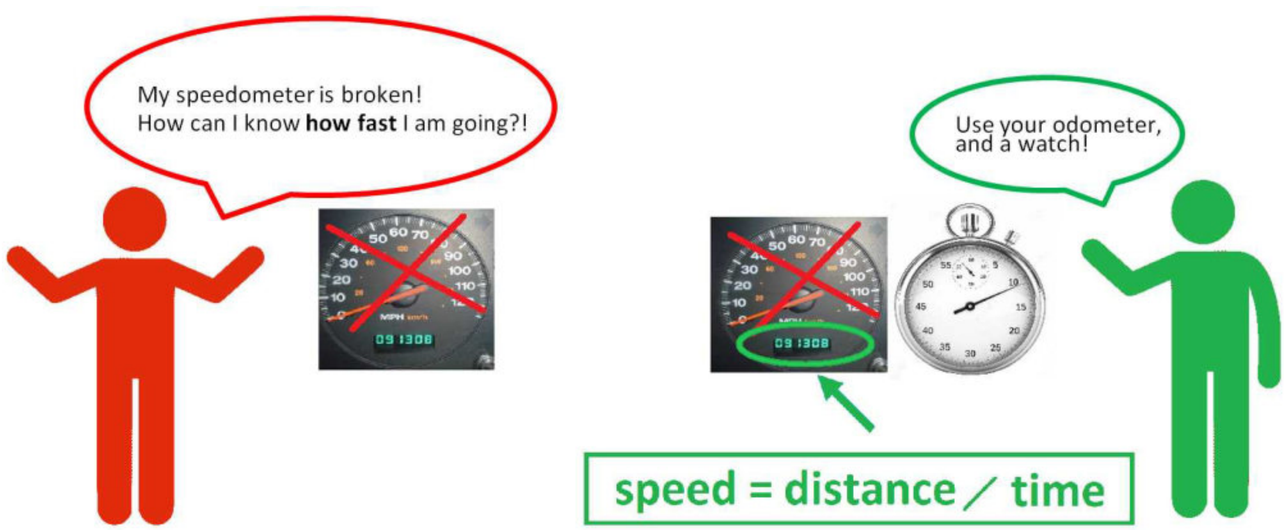
Chapter 1: Limits of sequences

Contents

1.1. What is calculus about?

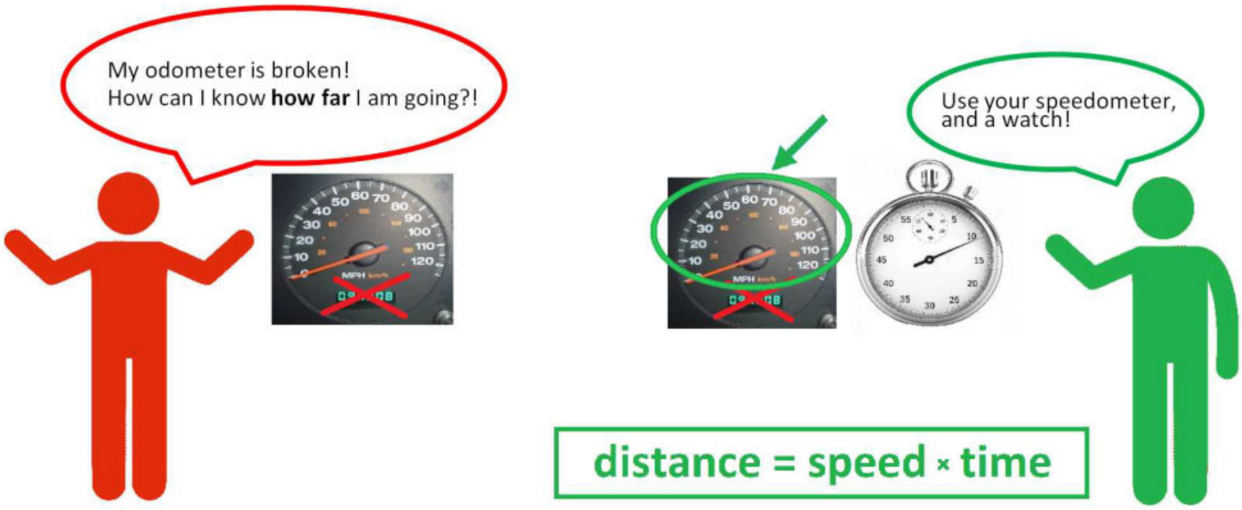
One of the main entry ways to calculus is the study of *motion*.

We present the idea of calculus in these two related pictures. First, we derive the speed from the distance that we have covered:



Beyond this conceivable situation, this formula is the definition of speed.

On the flip side, we derive the distance we have covered from the known velocity:



The two problems are solved, respectively, with the help of these two versions of the same elementary school formula:

$$\boxed{\text{speed} = \text{distance} \,/\, \text{time} \qquad \text{and} \qquad \text{distance} = \text{speed} \times \text{time}}$$

We solve the equation for the distance or for the speed depending on what is known and what is unknown.

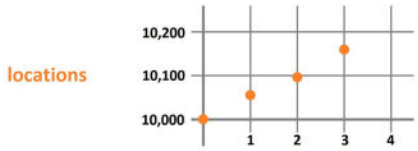
What takes this idea beyond elementary school is the possibility that *velocity varies* over time. The simplest case is when it varies incrementally.

Let's be more specific. We will face the two situations above but with more data collected and more information derived from it.

First, imagine that our speedometer is broken. What do we do if we want to estimate how fast we are driving during our trip? We look at the odometer *several* times – say, every hour on the hour – during the trip and record the mileage on a piece of paper. The list of our consecutive *locations* might look like this:

- initial reading: $10,000$ miles
- after the first hour: $10,055$ miles
- after the second hour: $10,095$ miles
- after the third hour: $10,155$ miles
- etc.

We can plot – as an illustration – the locations against time:



But what do we know about what the speed has been? We write a quick formula:

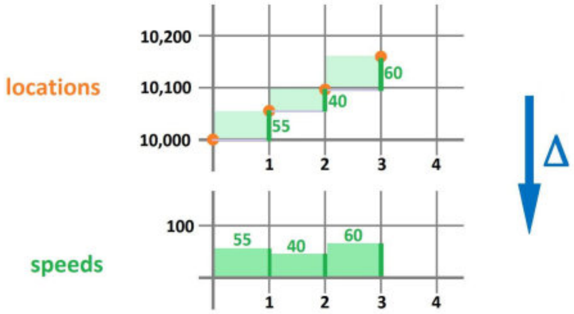
$$\boxed{\text{speed} = \frac{\text{distance}}{\text{time}} = \frac{\text{current location} - \text{last location}}{1}}$$

The time interval was chosen to be 1 hour, so all we need is to find the distance covered during each of these one-hour periods, by *subtraction*:

- distance covered during the first hour: $10,055 - 10,000 = 55$ miles; speed 55 miles an hour
- distance covered during the second hour: $10,095 - 10,055 = 40$ miles; speed 40 miles an hour
- distance covered during the third hour: $10,155 - 10,095 = 60$ miles; speed 60 miles an hour

- etc.

We see below how these new numbers appear as the blocks that make up each step of our last plot (top):



We then lower these blocks to the bottom to create a new plot (bottom).

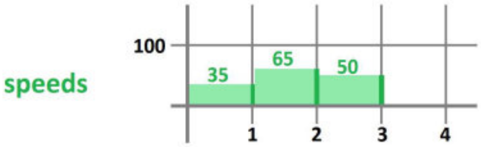
As you can see, we illustrate the new data in such a way as to suggest that the speed remains *constant* during each of these hour-long periods.

The problem is solved! We have established that the speed has been – roughly – 55, 40, and 60 miles an hour during those three time intervals, respectively.

Now on the flip side: Imagine this time that it is the odometer that is broken. If we want to estimate how far we will have gone, we should look at the speedometer *several* times – say, every hour – during the trip and record its readings on a piece of paper. The result may look like this:

- during the first hour: 35 miles an hour
- during the second hour: 65 miles an hour
- during the third hour: 50 miles an hour
- etc.

Let's plot our speed against time to visualize what has happened:



Once again, we illustrate the data in such a way as to suggest that the speed remains *constant* during each of these hour-long periods.

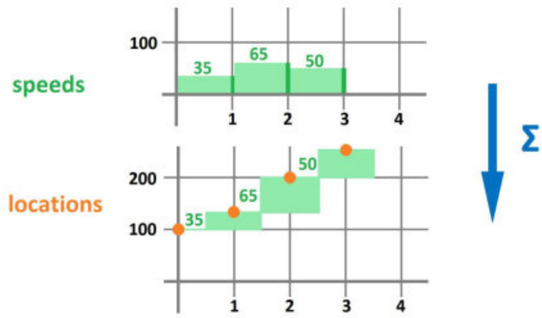
Now, what does this tell us about our location? We write a quick formula:

$$\boxed{\text{distance} = \text{speed} \times \text{time} = \text{speed} \times 1}$$

In contrast to the former problem, we need another bit of information. We must know the *starting point* of our trip, say, the 100-mile mark. The time interval was chosen to be 1 hour so that we need only to *add*, and keep adding, the speed at which – we assume – we drove during each of these one-hour periods:

- the location after the first hour: $100 + 35 = 135$-mile mark
- the location after the two hours: $135 + 65 = 200$-mile mark
- the location after the three hours: $200 + 50 = 250$-mile mark
- etc.

In order to illustrate this algebra, we plot the speeds as these blocks (top):



Then we use these blocks to make the consecutive steps of the staircase to show how high we have to climb (bottom).

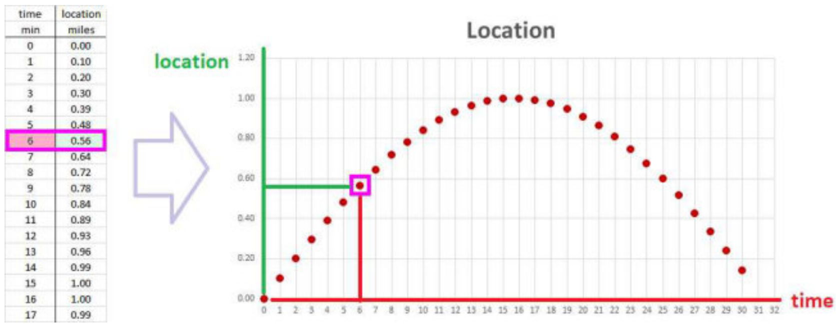
The problem is solved! We have established that we have progressed through the roughly 135-, 200-, and 250-mile marks during this time.

We next consider more complex examples of the relation between location and velocity. First, *from location to velocity.*

Suppose that this time we have a *sequence* of more than 30 data points (more is indicated by "..."); they are the locations of a moving object recorded every minute:

time	minutes	0	1	2	3	4	5	6	7	8	9	10	...
location	miles	0.00	0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	0.84	...

This data is also seen in the first two columns of the spreadsheet (left):



Every pair of numbers in the table is then plotted (right). The "scatter plot" that illustrates the data looks like a *curve*!

What has happened to the moving object can now be read from the graph. Just as in the last example, we concentrate on the vertical increment of the staircase:

These are the results:

1. The object was moving in the positive direction.
2. It was moving fairly fast but then started to slow down.
3. It stopped for a very short period.
4. Then it started to move in the opposite direction.
5. Then it started to speed up in that direction.

To understand how fast we move over these one-minute intervals, we compute the *differences* of locations for each pair of consecutive locations.

First, the table.

We use the data from the row of locations and subtract every two consecutive locations. This is how the first step is carried out:

time	min	0	1	...
location	miles	0.00	0.10	...
		$\searrow$	$\downarrow$	
difference			$0.10 - 0.00$	...
			$\Vert$	
velocity	miles/min		0.10	...

We compute this difference for each pair of consecutive locations and then place it in a row for the velocities that we added to the bottom of our table:

time	min	0	1	2	3	4	5	6	7	8	9	...
location	miles	0.00	0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	...
			$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	$\searrow$ $\downarrow$	...
velocity	miles/min		0.10	0.10	0.10	0.09	0.09	0.09	0.08	0.07	0.07	...

We have a new sequence!

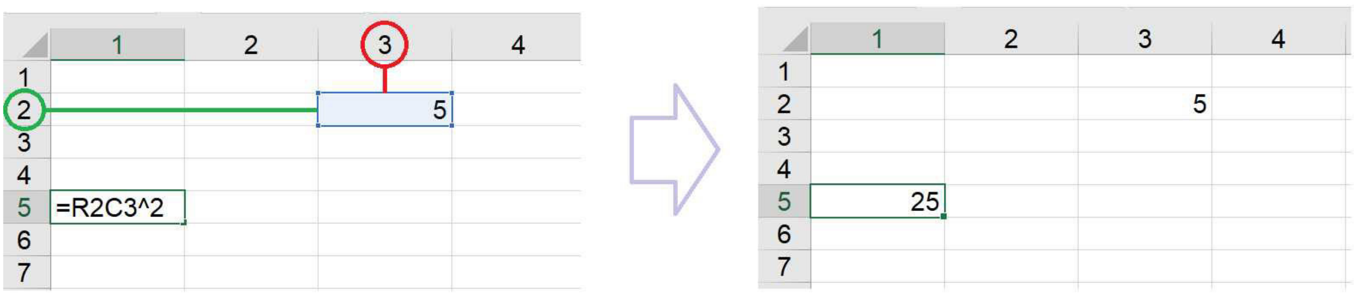
Practically, we'd rather use the computing capabilities of the spreadsheet.

Example 1.1.1: spreadsheet formulas

We use formulas to pull data from other cells. There are two ways. First, the “absolute" reference:

```
=R2C3^2
```

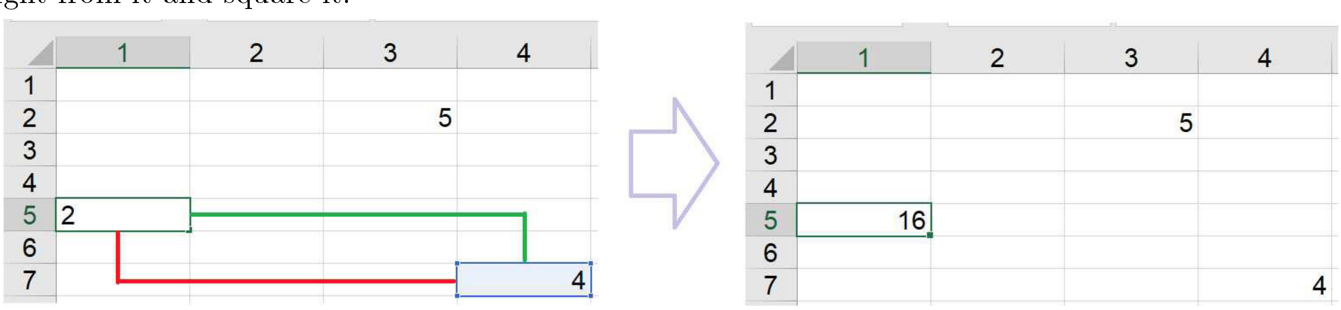
Any cell with this formula will take the value contained in the cell located at row 2 and column 3 and square it:



Second, the “relative" reference:

```
=R[2]C[3]^2
```

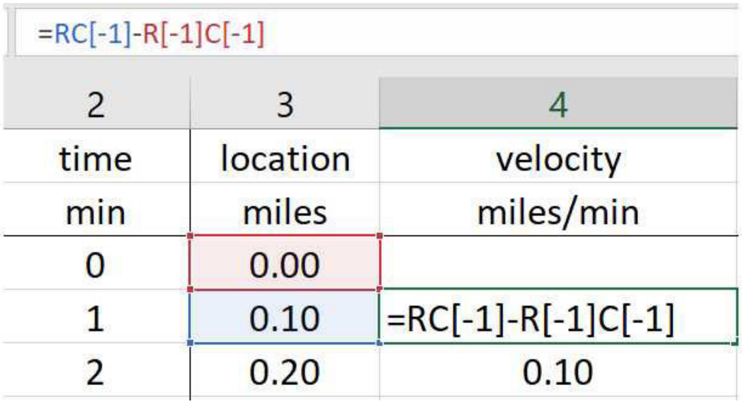
Any cell with this formula will take the value contained in the cell located 2 rows down and 3 columns right from it and square it:



We compute the differences by pulling data from the column of locations with the following formula:

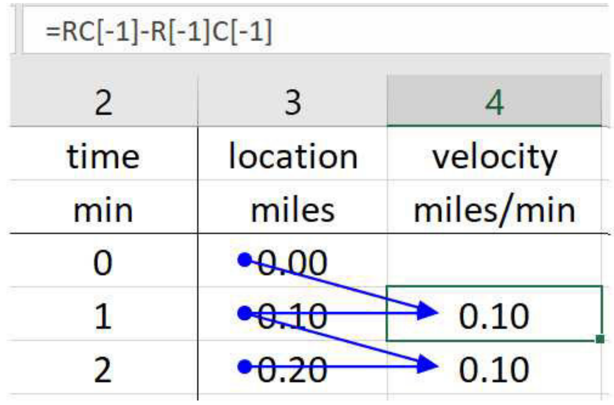
```
=RC[-1]-R[-1]C[-1]
```

Here, the two values come from the last column, `C[-1]`, same row, `R`, and last row, `R[-1]`. Below, you can see the two references in the formulas marked with red and blue (left) and the dependence shown with the arrows (right):



=RC[-1]-R[-1]C[-1]

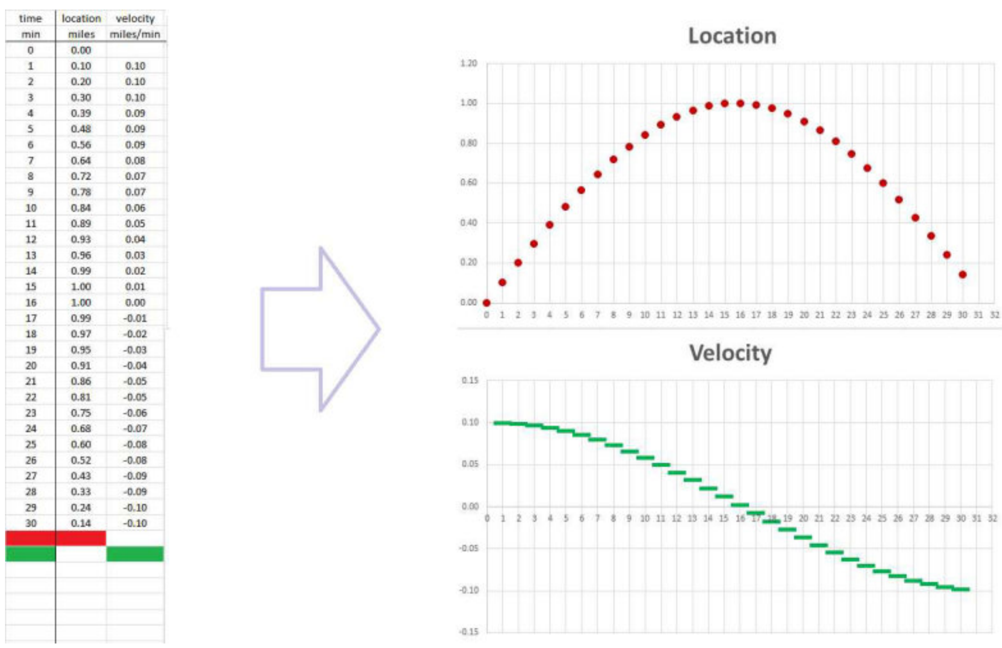
2	3	4
time	location	velocity
min	miles	miles/min
0	0.00	
1	0.10	=RC[-1]-R[-1]C[-1]
2	0.20	0.10



=RC[-1]-R[-1]C[-1]

2	3	4
time	location	velocity
min	miles	miles/min
0	0.00	
1	0.10	0.10
2	0.20	0.10

We place the result in a new column we created for the velocities:



time	location	velocity
min	miles	miles/min
0	0.00	
1	0.10	0.10
2	0.20	0.10
3	0.30	0.10
4	0.39	0.09
5	0.48	0.09
6	0.56	0.09
7	0.64	0.08
8	0.72	0.07
9	0.78	0.07
10	0.84	0.06
11	0.89	0.05
12	0.93	0.04
13	0.96	0.03
14	0.99	0.02
15	1.00	0.01
16	1.00	0.00
17	0.99	-0.01
18	0.97	-0.02
19	0.95	-0.03
20	0.91	-0.04
21	0.86	-0.05
22	0.81	-0.05
23	0.75	-0.06
24	0.68	-0.07
25	0.60	-0.08
26	0.52	-0.08
27	0.43	-0.09
28	0.33	-0.09
29	0.24	-0.10
30	0.14	-0.10

This new data is illustrated with the second scatter plot. To emphasize the fact that the velocity data, unlike the location, is referring to time intervals rather than time instances, we plot it with horizontal segments. In fact, the data table can be rearranged as follows to make this point clearer:

time	0		1		2		3		4		...
location	0.00	–	0.10	–	0.20	–	0.30	–	.39	–	...
velocity	·	0.10	·	0.10	·	0.10	·	0.09	·	0.09	...

What has happened to the moving object can now be easily read from the second graph. These are the five stages:

1. The velocity was positive initially (it was moving in the positive direction).
2. The velocity was fairly high (it was moving fairly fast) but then it started to decline (slow down).
3. The velocity was zero (it stopped) for a very short period.
4. Then the velocity became negative (it started to move in the opposite direction).
5. And then the velocity started to become more negative (it started to speed up in that direction).

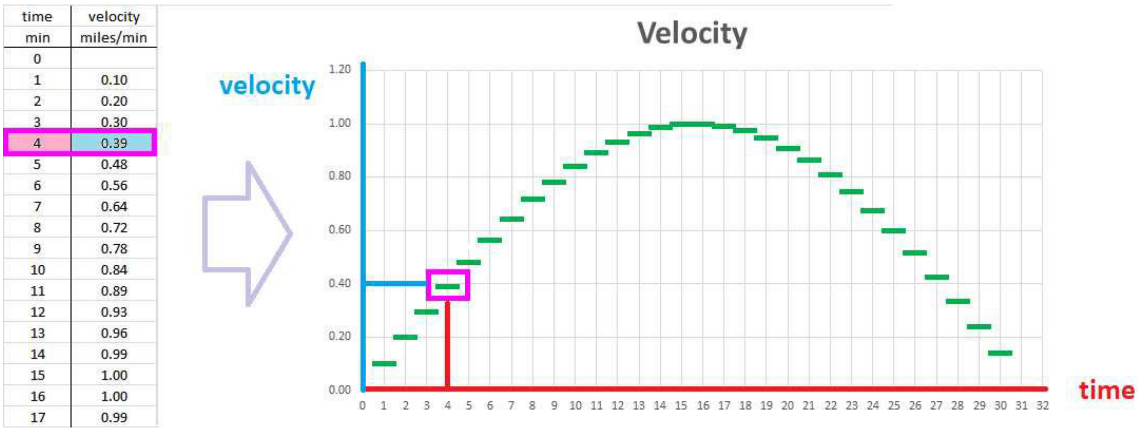
Thus, the latter set of data succinctly records some important facts about the dynamics of the former.

Now, *from velocity to location.*

Again, we consider a sequence of 30 data points. These numbers are the values of the velocity of an object recorded every minute:

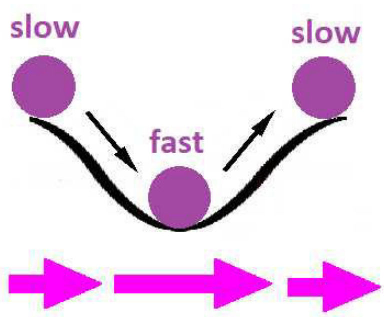
time	minutes	0	1	2	3	4	5	6	7	8	9	10	...
velocity	miles/hour		0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	0.84	...

This data is also seen in the first two columns of the spreadsheet plotted one bar at a time:



The data is furthermore illustrated as a scatter plot on the right. Again, we emphasize the fact that the velocity data is referring to time intervals by plotting its values with horizontal bars.

The data may be describing the horizontal speed of a ball rolling through a trough:



To find out where we are at the end of each of these one-minute intervals, we compute the *adding* the velocities one at a time. This is how the first step is carried out, under the assumption that the initial location is 0:

time	min	0	1	...
velocity	miles		0.10	...
			$\downarrow$	
sum		0.00+	0.10	...
		$\uparrow$	$\vert\vert$	
location	miles/min	0.00	0.10	...

We place this data in a new row added to the bottom of our table:

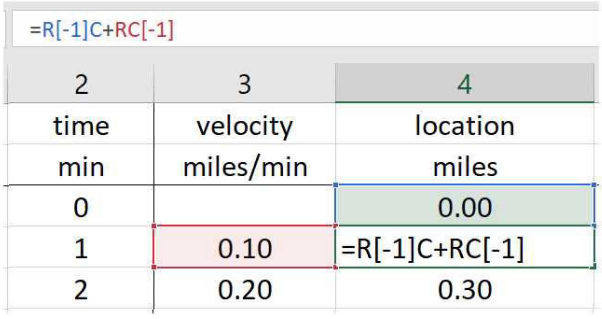
time	min	0	1	2	3	4	5	6	7	8	...
velocity	miles		0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	...
			$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	$\downarrow$	...
location	miles/min	0.00 $\to$	0.10 $\to$	0.30 $\to$	0.59 $\to$	0.98 $\to$	1.46 $\to$	2.03 $\to$	2.67 $\to$	3.39 $\to$	...

We have a new sequence!

Practically, we use the spreadsheet. We compute the sums by pulling the data from the column of velocities using the following formula:

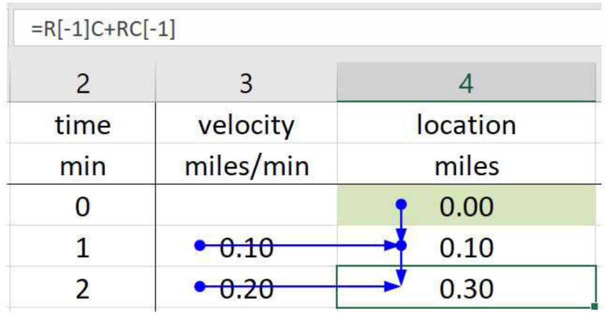
```
=R[-1]C+RC[-1]
```

Here, the two values come from the same, `C`, or last, `C[-1]`, column and the same, `R`, and last, `R[-1]`, row, as follows:



=R[-1]C+RC[-1]

2	3	4
time	velocity	location
min	miles/min	miles
0		0.00
1	0.10	=R[-1]C+RC[-1]
2	0.20	0.30

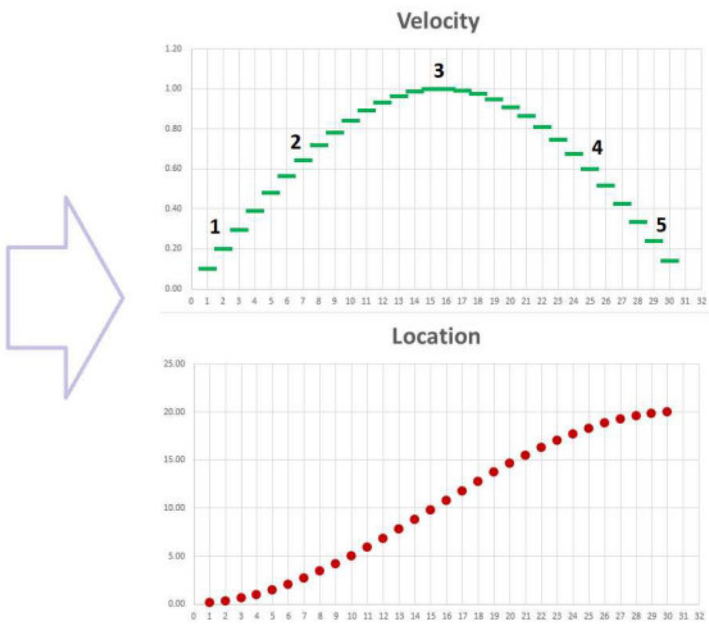


=R[-1]C+RC[-1]

2	3	4
time	velocity	location
min	miles/min	miles
0		0.00
1	0.10	0.10
2	0.20	0.30

We place the result in a new column for locations:

time	velocity	location
min	miles/min	miles
0		0.00
1	0.10	0.10
2	0.20	0.30
3	0.30	0.59
4	0.39	0.98
5	0.48	1.46
6	0.56	2.03
7	0.64	2.67
8	0.72	3.39
9	0.78	4.17
10	0.84	5.01
11	0.89	5.91
12	0.93	6.84
13	0.96	7.80
14	0.99	8.79
15	1.00	9.78
16	1.00	10.78
17	0.99	11.77
18	0.97	12.75
19	0.95	13.70
20	0.91	14.60
21	0.86	15.47
22	0.81	16.28
23	0.75	17.02
24	0.68	17.70
25	0.60	18.30
26	0.52	18.81
27	0.43	19.24
28	0.33	19.57
29	0.24	19.81
30	0.14	19.95



The data is also illustrated as the second scatter plot on the right.

What has happened to the moving object can now be easily read from the first or the second plot. These are the five stages:

1. The velocity is positive and low.
2. The velocity is positive and high.
3. The velocity is the highest.
4. The velocity is positive and high.
5. The velocity is positive and low.

We, again, rearrange the data table to make the difference between the two types of data clearer:

time	0		1		2		3		4		...
velocity	·	0.00	·	0.10	·	0.20	·	0.30	·	0.39	...
location	0.00	–	0.10	–	0.30	–	0.59	–	.98	–	...

Thus, as the former data set records some facts about the dynamics of the latter, we are able to use this information to recover the latter.

This is what we have discovered:

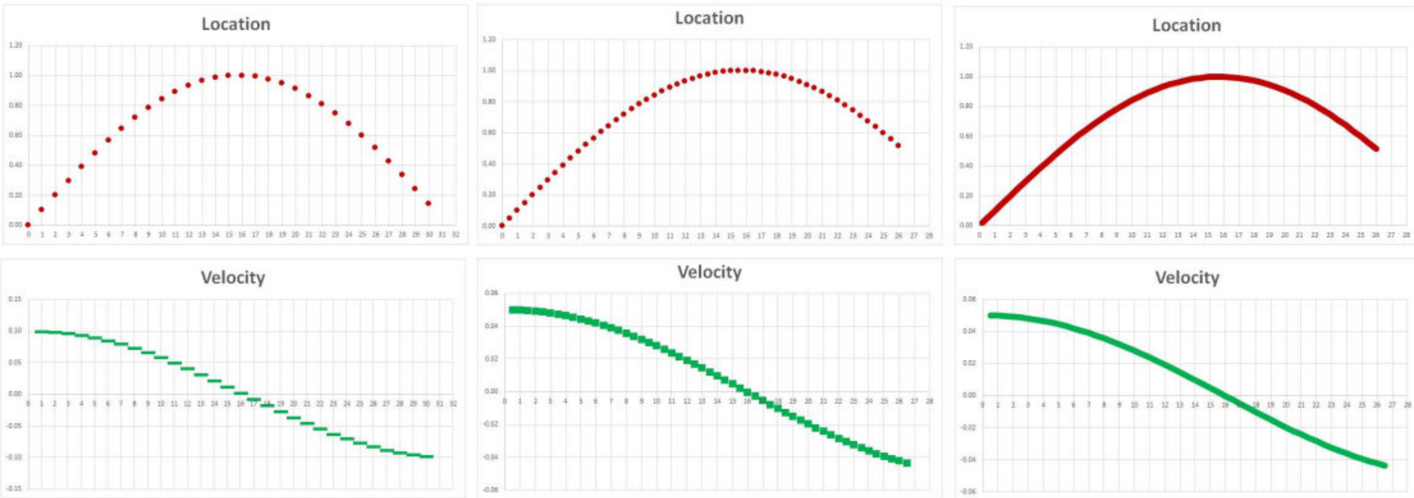
▶ *We can tell the velocity from the location and, conversely, the location from the velocity.*

Is this it though?

Motion is a *continuous* phenomenon. Can we understand it with the above approach?

If it is known that our data is just a snapshot of a "continuous" process, we may be able to collect more information in order to make this representation better. We, for example, may look at the odometer every

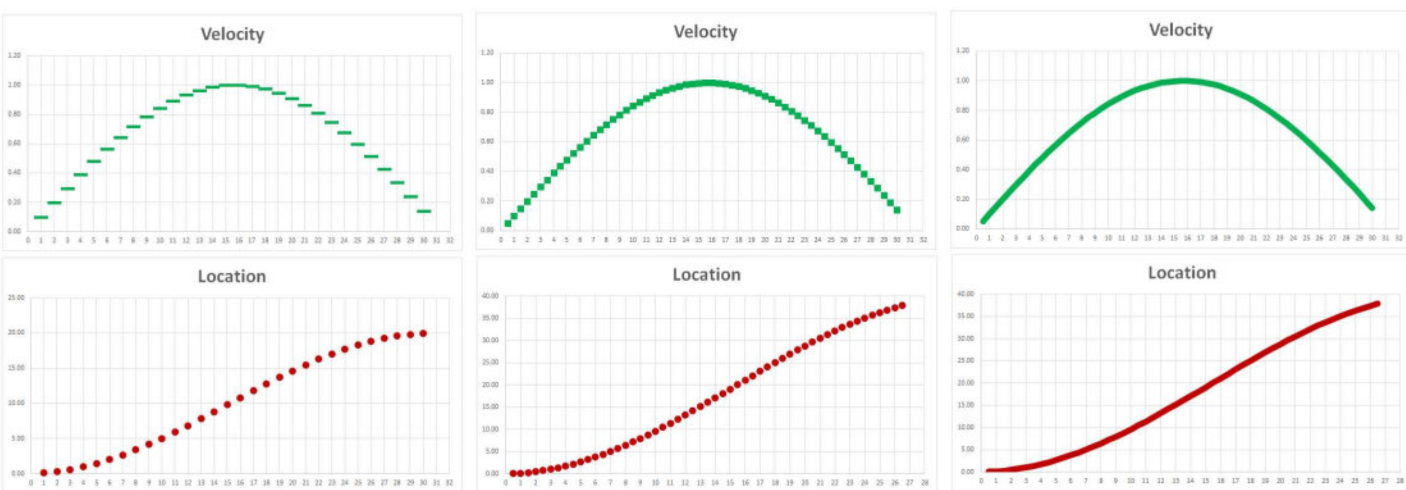
minute, or every second, etc., instead of every hour. The infinite divisibility of the real line (seen in Volume 1, Chapter 1PC-1) allows us to produce sets of points on the plane with denser and denser patterns: We make the time intervals smaller and smaller and insert more and more inputs. When there are enough of them, the points start to form a curve:



Both location and velocity are changing continuously!

We imagine that at the end of this process we will have an actual curve. This is not the kind of curve that is made of marbles placed close together, but a rope. What happens "at the end" is studied in Chapter 2.

Thus, this main idea of calculus is to *derive* velocity from location (from the top row to the bottom) and location from velocity (from the bottom row to the top):

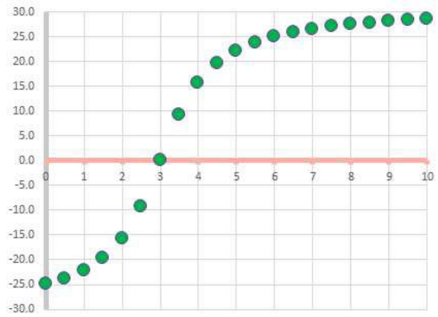


It is a special challenge to solve this problem for the *curves*. It is addressed in Chapter 3.

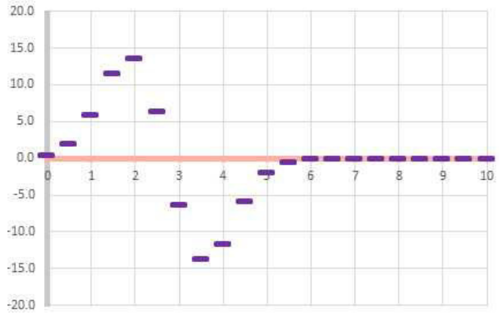
We, however, set these ideas aside for now and turn to something more immediate:

► What can we say about the *long-term trends* of the location and the velocity?

Let's take another look at the first chart above and the sequences plotted. Is the future *predictable*? The object is moving in the negative direction (top row), but we can't tell if this will continue. However, it is slowing down (bottom row), and it is conceivable that it will stop eventually. The pattern is just as clear in the second chart above or in this chart:



The motion shown in the next one has definitely stopped:



This and related issues (limited to sequences) will be addressed in this chapter. But first a review.

1.2. Infinite sequences and their long-term trends

We formalize the way we represent sequences of numbers such as the ones we saw in the last section:

time	minutes	0	1	2	3	4	5	6	7	8	9	10	...
location	miles	0.00	0.10	0.20	0.30	0.39	0.48	0.56	0.64	0.72	0.78	0.84	...

We first give a sequence a name, say, a, and then assign a specific variation of this name to each term of the sequence:

Indices of sequence

index:	n	1	2	3	4	5	6	7	...
term:	a_n	a_1	a_2	a_3	a_4	a_5	a_6	a_7	...

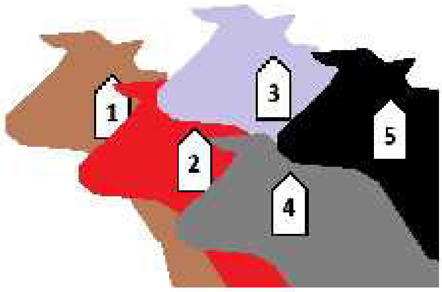
The *name* of a sequence is a letter, while the subscript called the *index* indicates the place of the term within the sequence. It reads "a sub 1", "a sub 2", etc.

This is what the notation means:

Index of a term

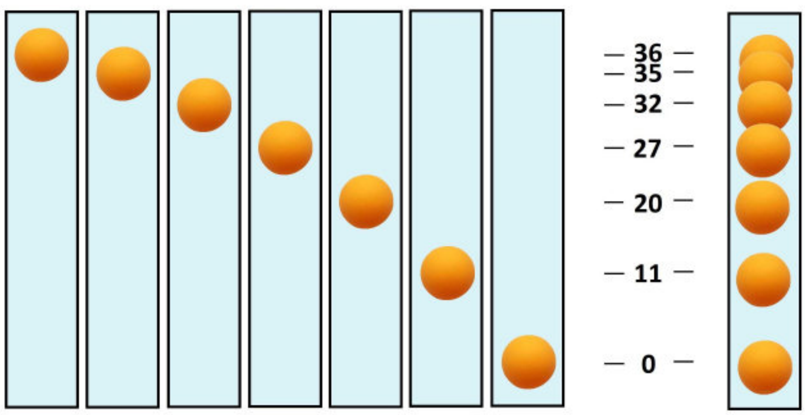
$$\underset{\substack{\uparrow\\ \text{name}}}{a}\ \overset{\substack{\text{index}\\ \downarrow}}{n}$$

Indices serve as *tags*:



Example 1.2.1: falling ball

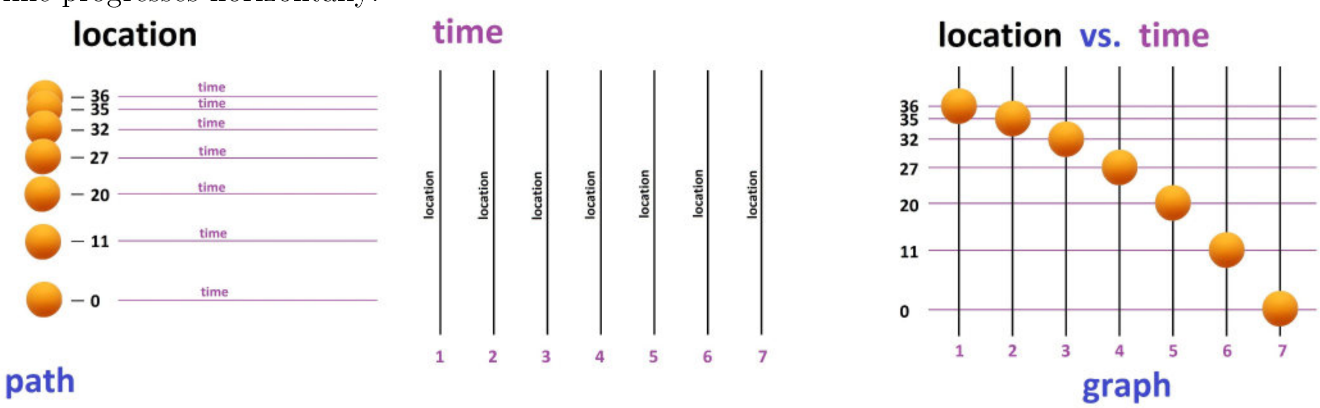
We watch a ping-pong ball falling down and record – at equal intervals – how high it is. The result is an ever-expanding string, a sequence, of numbers. If the frames of the video are combined into one image, it will look something like this:



We have a *list*:

$$36,\ 35,\ 32,\ 27,\ 20,\ 11,\ 0,\ ...$$

We bring them back together in one rectangular plot so that the location varies vertically while the time progresses horizontally:



The plot is called the *graph* of the sequence.

As far as the data is concerned, we have a list of *pairs*, time and location, arranged in a table:

moment	height
1	36
2	35
3	32
4	27
5	20
6	11
7	0
...	...

or

moment:	1	2	3	4	5	6	7	...
height:	36	35	32	27	20	11	0	...

In our example, we name the sequence h for "height". Then the above *table* take this form:

moment:	1	2	3	4	5	6	7	...
height:	h_1	h_2	h_3	h_4	h_5	h_6	h_7	...
	||	||	||	||	||	||	||	...
height:	36	35	32	27	20	11	0	...

When abbreviated, it takes the form of this *list*:

$$h_1 = 36,\ h_2 = 35,\ h_3 = 32,\ h_4 = 27,\ h_5 = 20,\ h_6 = 11,\ h_7 = 0,\ ....$$

So, we use the following notation:

$$a_1 = 1,\ a_2 = 1/2,\ a_3 = 1/3,\ a_4 = 1/4,\ ...$$

where a is the *name* of the sequence and adding a subscript indicates which term of the sequence we are facing.

We will study *infinite* sequences.

Such a sequence can be examined in terms of its long-term, or better yet infinite, *trend.* The idea is simple:

1. The sequence	1	1/2	1/3	1/4	1/5	...	tends toward 0.
2. The sequence	.9	.99	.999	.9999	.99999	...	tends toward 1.
3. The sequence	1	2	3	4	5	...	tends toward $+\infty$.
4. The sequence	0	1	0	1	0	...	doesn't tend to anything.

So, an infinite sequence of numbers will sometimes be "accumulating" around a single number, its "limit" (#1 and #2). It's as if the gap between the bouncing ball and the ground becomes invisible. This isn't always the case: The sequence might "run away" (#3) or bounce around without slowing down (#4).

Every function $y = f(x)$ with an appropriate domain creates a sequence via a simple substitution:

$$a_n = f(n)\,.$$

A function defined on a ray in the set of integers, $\{p, p+1, ...\}$, is called an infinite sequence, or simply sequence, and it is typically given by its formula:

$$a_n = 1/n, \quad n = 1, 2, 3, ...$$

In addition to tables and formulas, a sequence is commonly defined by computing its terms in a *consecutive manner*, one at a time.

Definition 1.2.2: recursive sequence

We say that a sequence is *recursive* when its next term is found from the current term by a specified formula, i.e., a_n determines a_{n+1}.

Example 1.2.3: bank account

A person starts to deposit \$20 every month in his bank account that already contains \$1000. In other words, we have:

$$\text{next} \ = \ \text{last} \ + 20\,,$$

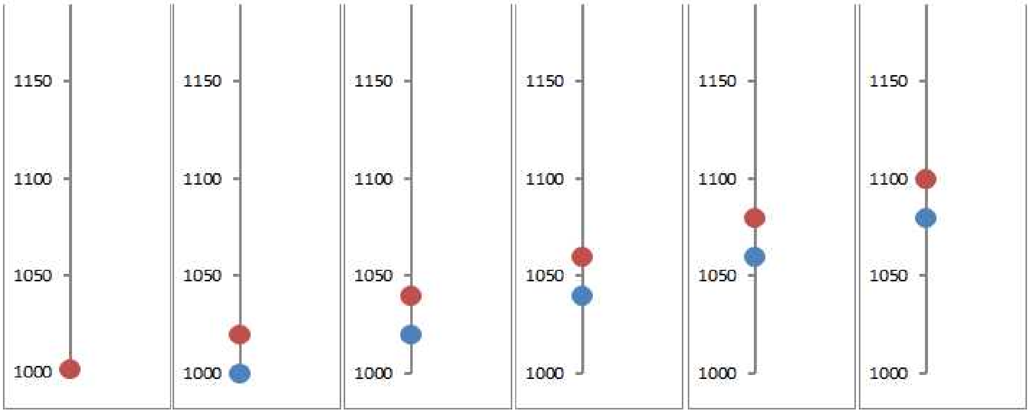
or with a *recursive formula*:

$$a_{n+1} = a_n + 20\,,$$

or for the spreadsheet:

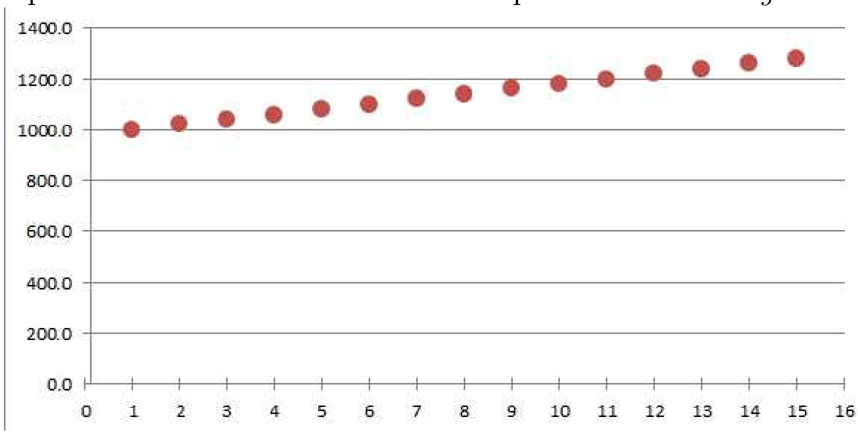
```
=R[-1]C+20
```

Below, the current amount is shown in blue, and the next – computed from the current – is shown in red:



Plotting several terms of the sequence at once confirms that the sequence is *increasing*:

time	amount
1	1000.00
2	1020.00
3	1040.00
4	1060.00
5	1080.00
6	1100.00
7	1120.00
8	1140.00
9	1160.00
10	1180.00
11	1200.00
12	1220.00
13	1240.00
14	1260.00
15	1280.00



It also looks like a straight line.

Another person deposits \$1000 in his bank account that pays 1% APR compounded annually. In other words, we have:

$$\text{next} \;=\; \text{last} \;\cdot 1.01\,,$$

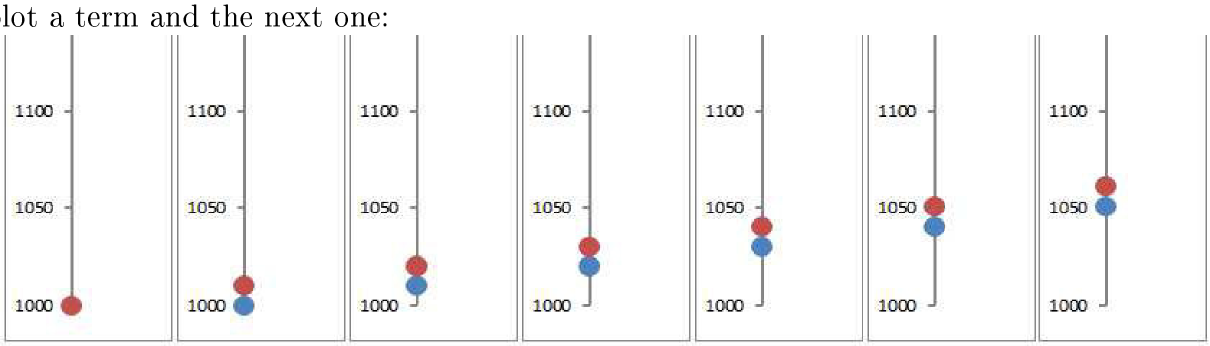
or with a *recursive formula*:

$$b_{n+1} = b_n \cdot 1.01\,,$$

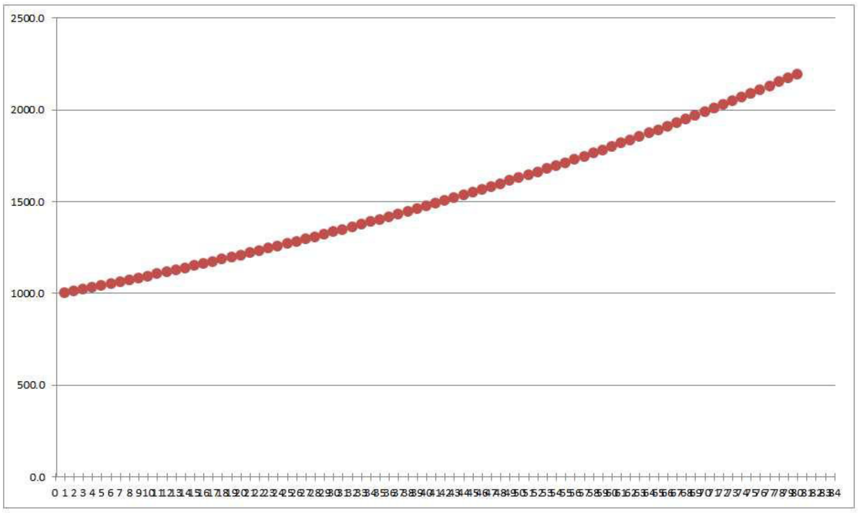
or for the spreadsheet:

```
=R[-1]C*1.01
```

We plot a term and the next one:



Only after repeating the step 100 times can one see that this isn't just a straight line:



What if we deposit money to our bank account *and* receive interest? The recursive formula is simple; for example:

$$c_{n+1} = c_n \cdot 1.05 + 2000\,.$$

Here, the interest is 5% with a \$2000 annual deposit.

Exercise 1.2.4

Show that $\dfrac{n}{n+1}$ is an increasing sequence. What kind of sequence is $\dfrac{n+1}{n}$? Give examples of increasing and decreasing sequences.

The following two concepts will be routinely used.

Definition 1.2.5: arithmetic progression

A sequence defined (recursively) by the formula:

$$\boxed{a_{n+1} = a_n + b}$$

is called an *arithmetic progression* with b as its *increment*.

Definition 1.2.6: geometric progression

A sequence defined (recursively) by the formula:

$$\boxed{b_{n+1} = b_n \cdot r}$$

with $r \neq 0$, is called a *geometric progression* with r as its *ratio*. We say that this is:

- a *geometric growth* when $r > 1$, and
- a *geometric decay* when $r < 1$.

Alternatively, it is called an *exponential* growth and decay, respectively.

Theorem 1.2.7: Formula for Arithmetic Progression

The nth-term formula for an arithmetic progression with increment a (that starts with c) is

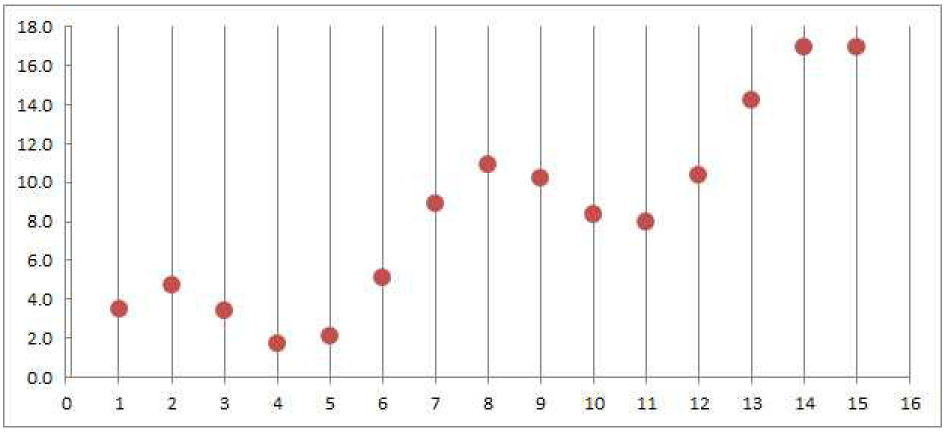
$$a_n = c + an\ , \quad n = 0, 1, 2, 3, \ldots$$

Theorem 1.2.8: Formulas for Geometric Progressions

The nth-term formula for a geometric progression with ratio r (that starts with c) is

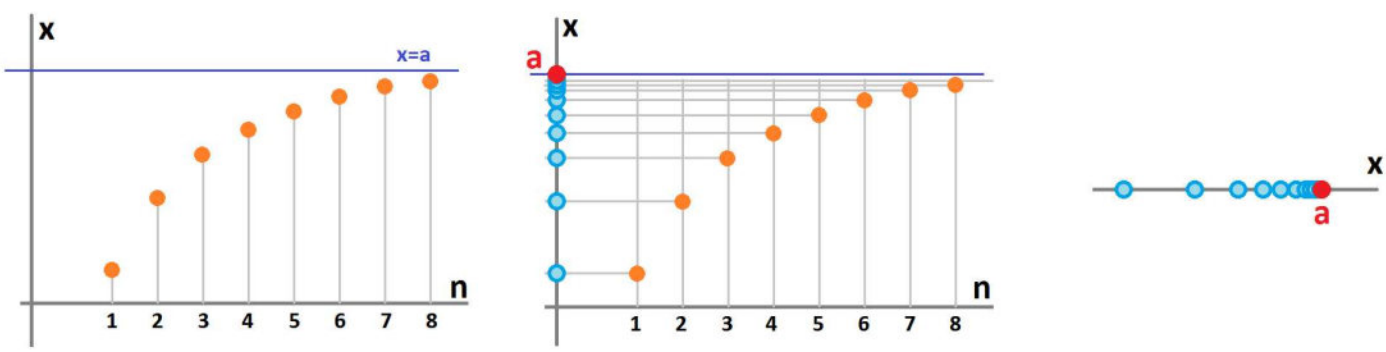
$$b_n = cr^n\ , \quad n = 0, 1, 2, 3, \dots$$

Sequences aren't usually this specific. We could visualize sequences as the graphs of functions:



However, we take a different approach in this chapter; we would like to apply, at a later time, what we have learned about sequences to our study of functions. This is why our visualizations of graphs of sequences will use a Cartesian coordinate system with these axes:

- the *horizontal axis* is the n-axis, and
- the *vertical axis* is the x-axis.



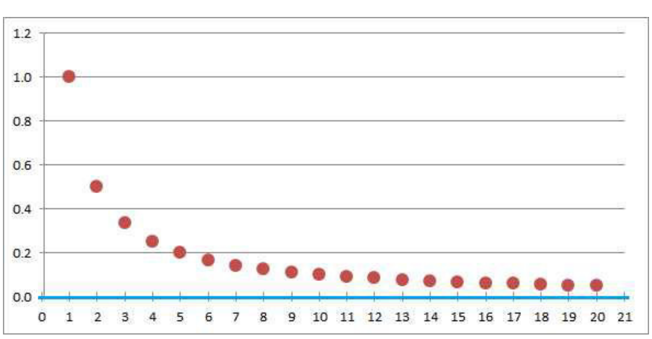
This approach allows us to have a more compact way to visualize sequences (right) as *sequences of locations* on the x-axis visited over an infinite period of time. The long-term trend becomes clear when the points stop visibly "moving".

Example 1.2.9: reciprocals

The go-to example is the sequence of the reciprocals:

$$x_n = \frac{1}{n}\ .$$

It appears to *tend to* 0.

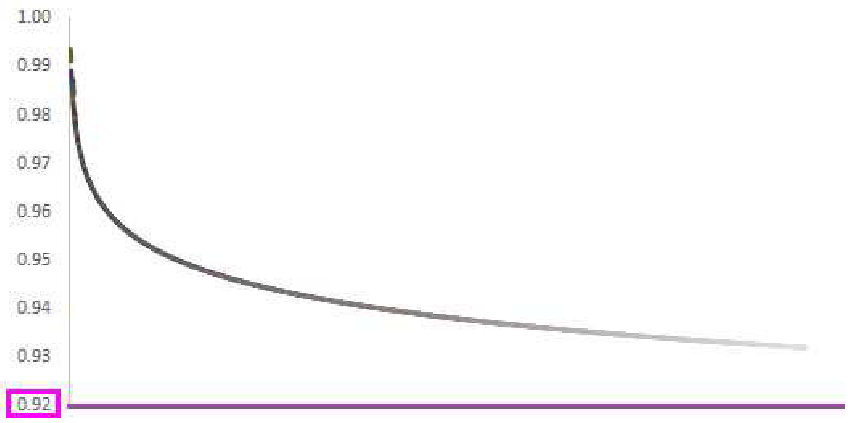


This fact is easy to confirm numerically:

$$x_n = 1.000,\ 0.500,\ 0.333,\ 0.250,\ 0.200,\ 0.167,\ 0.143,\ 0.125,\ 0.111,\ \ldots$$

Example 1.2.10: plotting

However, numerical analysis alone can't be used for discovering the value of the limit. Plotting the first 1000 terms of the sequence $x_n = n^{-.01}$ fails to suggest the true value of the limit:

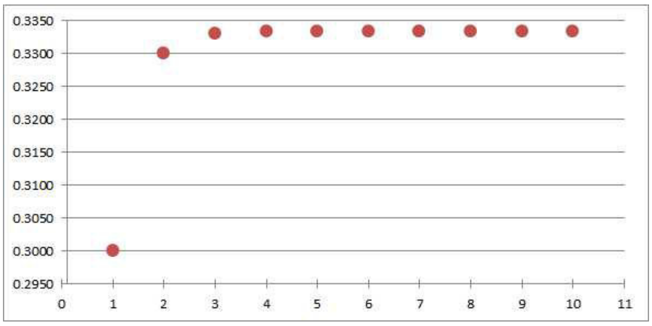


In fact, it is zero.

Example 1.2.11: decimals

Sequences are ubiquitous. For example, given a real number, we can easily construct a sequence that tends to that number – via its decimal approximations. For example,

$$x_n = 0.3,\ 0.33,\ 0.333,\ 0.3333,\ \ldots \text{ tends to } 1/3\,.$$



Example 1.2.12: alternating

When a sequence changes its signs all the time, we don't expect it to have a limit. An example is an "alternating sequence":

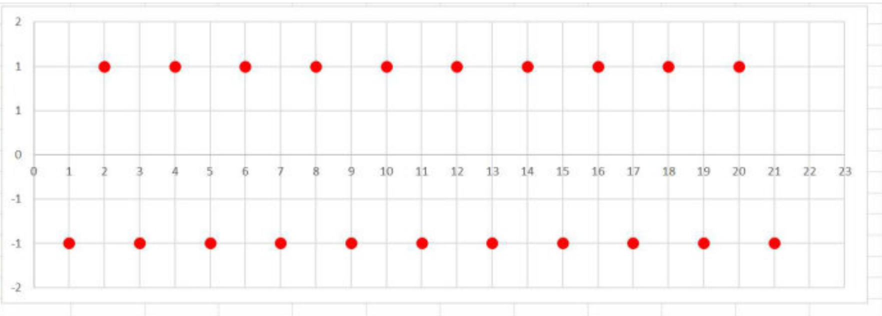
$$a_n = (-1)^n\,.$$

In other words, we have:

1. It's -1 when n is odd.
2. It's 1 when n is even.

The sequence alternates between these two numbers:

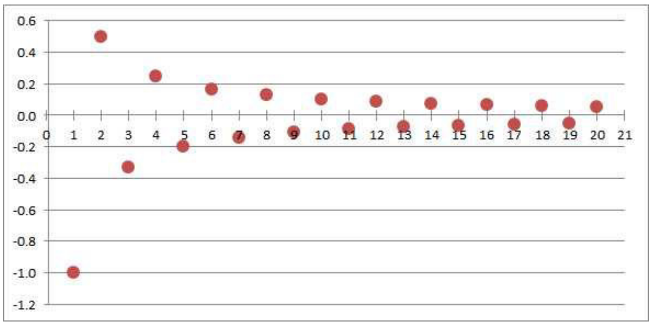
n	An
1	-1
2	1
3	-1
4	1
5	-1
6	1
7	-1
8	1
9	-1
10	1
11	-1
12	1
13	-1
14	1



But what if the swing is diminishing, such as this "alternating reciprocals sequence":

$$b_n = (-1)^n \frac{1}{n}\ ?$$

The values approach the ultimate destination – from both sides:



Exercise 1.2.13

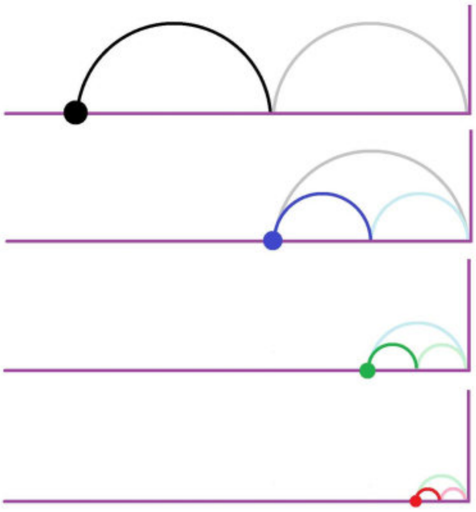
What can you say about the limit of an integer-valued sequence?

Example 1.2.14: Zeno's paradox

Consider a simple scenario: As you walk toward a wall, you can never reach it because:

1. Once you've covered half the distance, there is still distance left.
2. Once you've covered half of that distance, there is still more left.
3. And so on.

We have a sequence:



We mark these steps and do realize that there are infinitely many of them to be taken! How is this possible?

A short answer is: Just as an interval of *space* can (and is) split into infinitely many pieces, so can an interval of *time*. (The issue is addressed in Volume 3, Chapter 3IC-5.)

1.3. The definition of limit

Calculus, in large part, is the study of how to properly handle *infinity*.

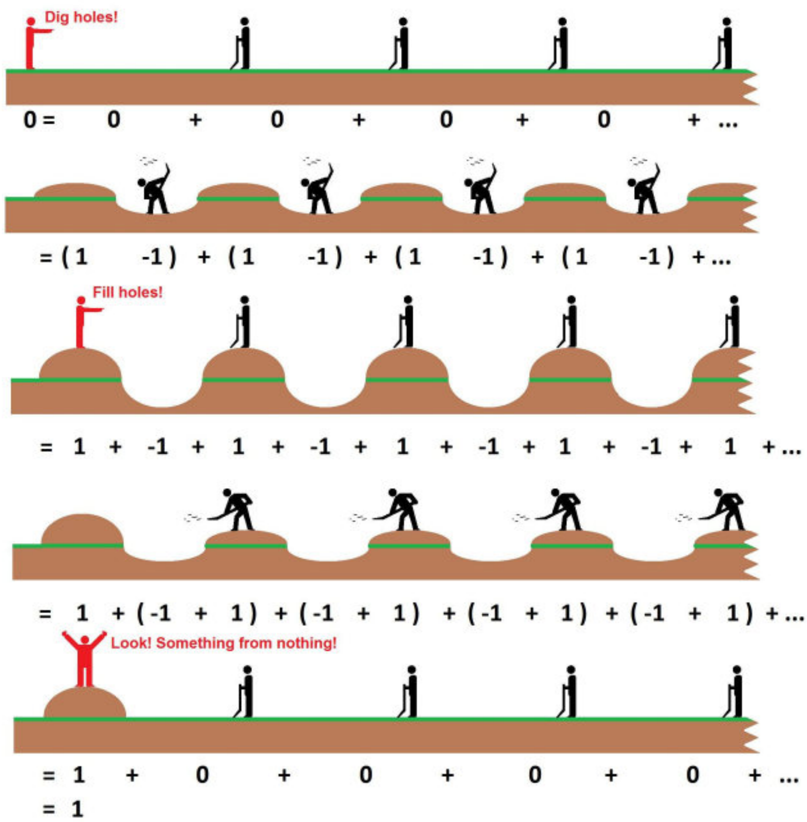
Example 1.3.1: something from nothing?

Let's examine this seemingly legitimate computation:

$$\begin{array}{llllllllllllll}
(1) & 0 & \overset{?}{=} & 0 & & +0 & & +0 & & +0 & & +\ldots & & \text{We are adding 0s.}\\
(2) & & \overset{?}{=} & (1 & -1) & +(1 & -1) & +(1 & -1) & +(1 & -1) & +\ldots & & \text{We use } 0=1-1\,.\\
(3) & & \overset{?}{=} & 1 & -1 & +1 & -1 & +1 & -1 & +1 & -1 & +\ldots & & \\
(4) & & \overset{?}{=} & 1 & +(-1 & +1) & +(-1 & +1) & +(-1 & +1) & +(-1 & +1) & \ldots & \text{We use } -1+1=0\,.\\
(5) & & \overset{?}{=} & 1 & +0 & & +0 & & +0 & & +0 & & +\ldots & \text{We are adding 0s.}\\
(6) & & \overset{?}{=} & 1\,. & & & & & & & & & &
\end{array}$$

That's impossible! What, $0 = 1$? How did this happen?

Consider a little story below:



The numbers refer to the amount of soil taken out, and one can say that we got *something from nothing*!

Our mistake was to be too casual about carrying out *infinitely many* algebraic operations.

Exercise 1.3.2

Which of the "=" signs above is incorrect? Hint: Think of this computation as a process.

The question has been:

- ▶ With this sequence, what number do its values approach?

We can also turn this around:

- ▶ With this number, the values of what sequence approach it?

So, the limit is a number and the sequence *approximates* this number:

(1)	The sequence	1	1/2	1/3	1/4	1/5	...	approximates 0 .
(2)	The sequence	.9	.99	.999	.9999	.99999	...	approximates 1 .
(3)	The sequence	1.	1.1	1.01	1.001	1.0001	...	approximates 1 .
(4)	The sequence	3.	3.1	3.14	3.141	3.1415	...	approximates π .
(5)	The sequence	1	2	3	4	5	...	approaches ∞ .
(6)	The sequence	0	1	0	1	0	...	doesn't approximate any number.

So, we can substitute the sequence for the number it approximates and do it with *any degree of accuracy*!

We use the following notation for the limit of a sequence:

Limit of sequence

$$a_n \to a$$

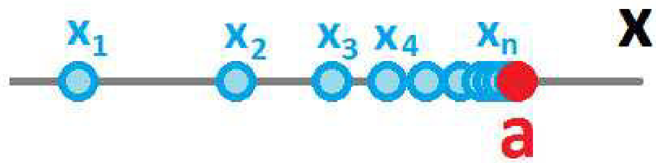
We can rewrite the above:

list								nth-term formula
(1)	1	1/2	1/3	1/4	1/5	...	$\to 0$	$1/n \to 0$
(2)	.9	.99	.999	.9999	.99999	...	$\to 1$	$1 - 10^{-n} \to 1$
(3)	1.	1.1	1.01	1.001	1.0001	...	$\to 1$	$1 + 10^{-n} \to 1$
(4)	3.	3.1	3.14	3.141	3.1415	...	$\to \pi$	
(5)	1	2	3	4	5	...	$\to +\infty$	$n \to +\infty$
(6)	0	1	0	1	0	...	$\to$ nothing	

Now, let's find the exact meaning of limit.

Example 1.3.3: trajectory

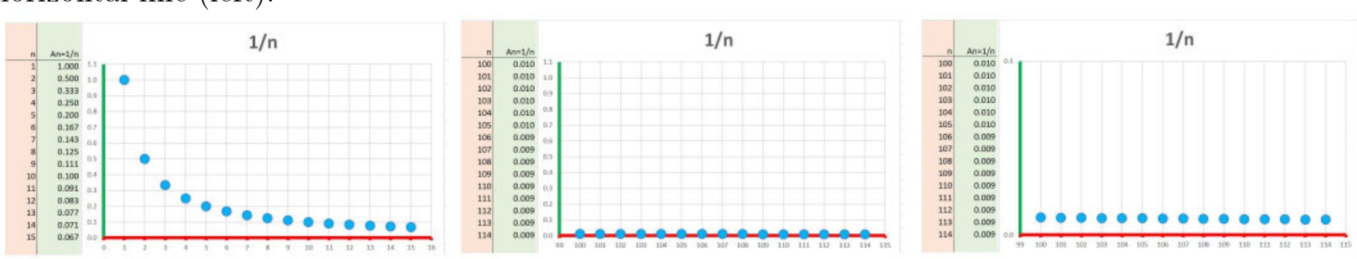
The values of a convergent sequence accumulate toward a particular number, just as in the example of a moving ball:



We can see that after sufficiently many steps, the terms of the sequence, a_n, become indistinguishable from the limit, a. It seems that, say, the 10th dot has merged with a.

Example 1.3.4: graph

The fact that the values of a sequence accumulate toward a particular number means that, geometrically, we will always see how the graph of the sequence ends up closer and closer to a particular horizontal line (left):



At the end, the dots seem to land on the x-axis (middle). However, the gap becomes visible as we zoom in (right).

Example 1.3.5: numerical data

Let's now look at this "process" numerically. Suppose we have a specific sequence:

$$a_n = 1/n^2.$$

We compute a few dozens of its values and then ask: What does it mean that it approaches $a = 0$?

First, how long does it take to get within .1 from $a = 0$? Look up in the table of values: It takes 4 steps (red):

n	1	2	3	4	5	6	7	8	9	10
An	1.000000	0.250000	0.111111	0.062500	0.040000	0.027778	0.020408	0.015625	0.012346	0.010000

<.1

11	12	13	14	15	16	17	18	19	20
0.008264	0.006944	0.005917	0.005102	0.004444	0.003906	0.003460	0.003086	0.002770	0.002500

<.01

21	22	23	24	25	26	27	28	29	30
0.002268	0.002066	0.001890	0.001736	0.001600	0.001479	0.001372	0.001276	0.001189	0.001111

31	32	33	34
0.001041	0.000977	0.000918	0.000865

<.001

We just follow the list of numbers until they become less than the threshold .1.

Second, how long does it take to get within .01 from a? It takes 11 steps (green).

Third, how long does it take to get within .001 from a? It takes 32 steps (blue).

And so on. No matter how small a number I pick, *eventually* a_n will be that close to its limit.

Another interpretation of this analysis is in terms of *accuracy*. We understand the idea that $a_n = 1/n^2$ approaches $a = 0$ as follows: "The sequence approximates 0".

- First, what if we need the accuracy to be .1? Look up in the table of values: We need to compute 4 terms of the sequence or more.
- Second, what if we need the accuracy to be .01? At least 11 terms.
- Third, what if we need the accuracy to be .001? At least 32.
- And so on.

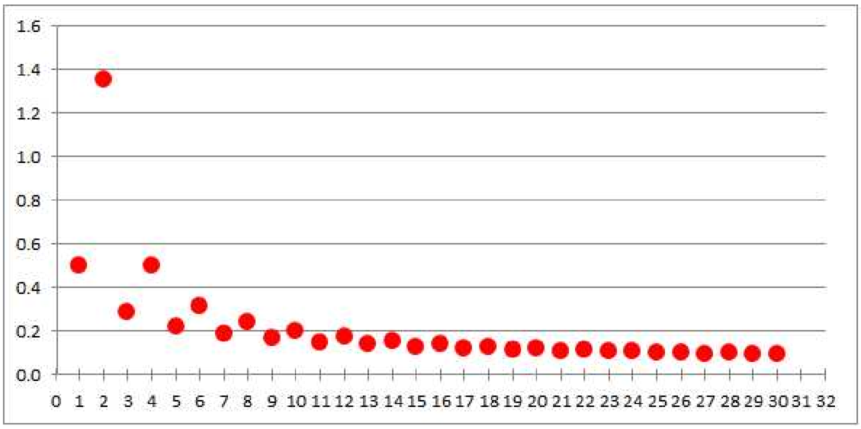
No matter how much accuracy I need, there is a way to accommodate this requirement by getting farther and farther into the sequence a_n.

Exercise 1.3.6

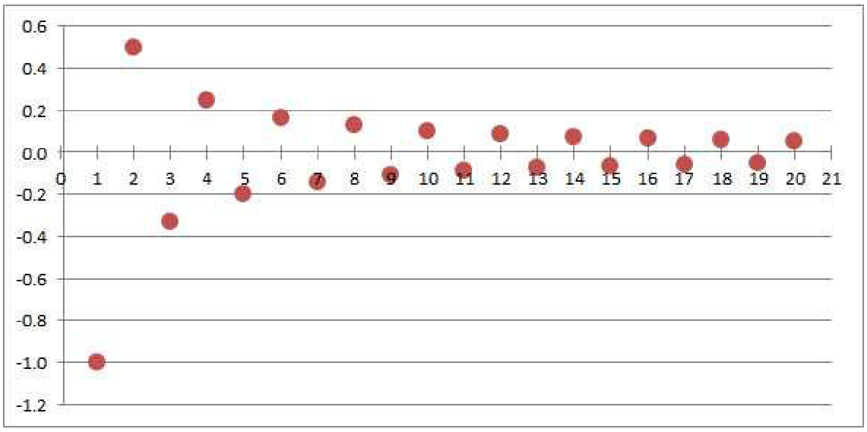
How long does it take to get within .0003 from $a = 0$?

Example 1.3.7: numerical data from graph

Unfortunately, not all sequences are as simple as that. They may approach their respective limits in a number of ways, as we have seen. They don't have to be monotone:



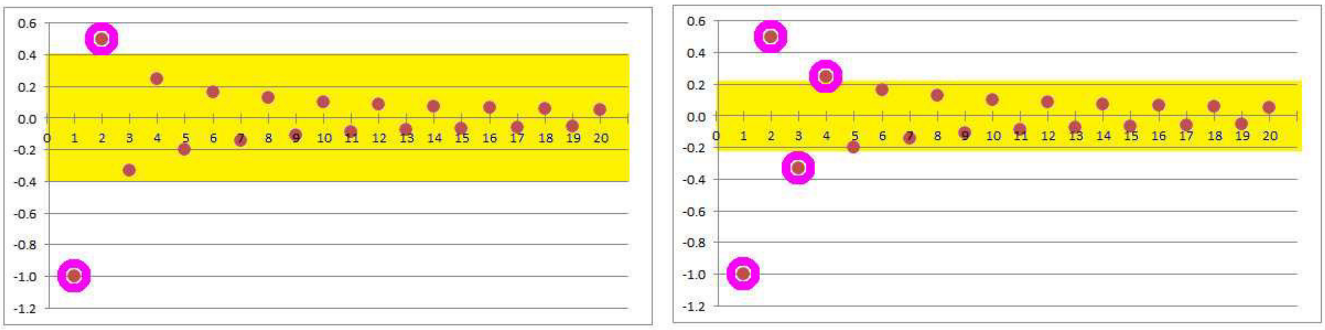
They might approach the limit from above and below at the same time:



Our questions can still be asked and answered:

- How long does it take to get within .4 from $a = 0$? Look up in the graph: It takes 4 steps to get within the band from $-.4$ to .4.
- How long does it take to get within .2 from $a = 0$? It takes 6 steps to get within the band from $-.2$ to .2.
- And so on.

This is how it is done:



Exercise 1.3.8

For the first graph in the example, how long does it take to get within .4, .2, .1 from $a = 0$?

Exercise 1.3.9

For the second graph in the example, how long does it take to get within .1 from $a = 0$?

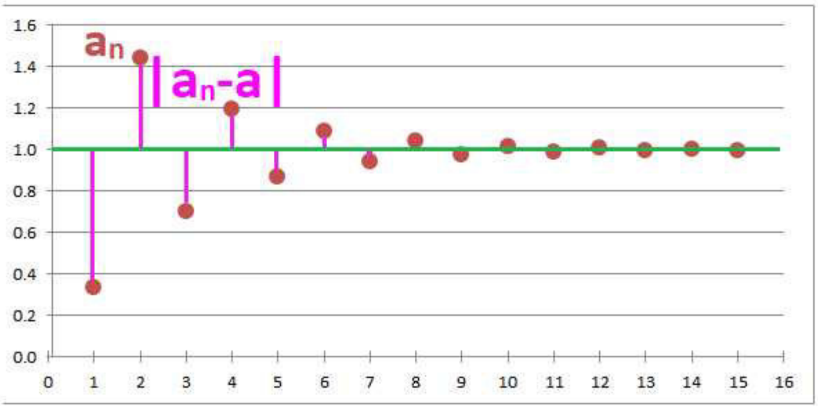
But what if the accuracy requirements keep increasing? It would be nice if we could solve the problem once and for all!

Below we rewrite what we want to say about the meaning of the limit in progressively more and more precise

terms:

	n	$y = a_n$
1.	As $n \to \infty$,	we have $y \to a$.
2.	As n approaches ∞,	y approaches a .
		We interpret "approach" in terms of distance.
3.	As n is getting larger and larger,	the distance from y to a goes 0 .
		The distance between y and a is $\lvert y-a \rvert$.
4.	By making n larger and larger,	we make $\lvert y-a \rvert$ as small as required.

The absolute values above are the distances from a_n to a, as shown below:



We make the definition even more precise:

	n	$y = a_n$
4.	By making n larger and larger,	we make $\lvert y-a \rvert$ as small as needed.
5.	By making n larger than some $N>0$,	we make $\lvert y-a \rvert$ smaller than any given $\varepsilon > 0$.

Algebraically, we see that for every measure of "closeness", call it ε, the function's values eventually become that close to the limit. In other words, ε is *the degree of required accuracy*.

Warning!

It is ε that comes first, then N.

Example 1.3.10: absolute accuracy

Let's consider the last iteration of our definition in light of the example above concerning the sequence:

$$a_n = 1/n^2.$$

These are the N's we got from those ε's:

By making n larger than some	$N > 0$,	we make $\lvert a_n - a \rvert$ smaller than any given	$\varepsilon > 0$.
	$N = 4$	$\longleftarrow$	$\varepsilon = .1$
	$N = 11$	$\longleftarrow$	$\varepsilon = .01$
	$N = 32$	$\longleftarrow$	$\varepsilon = .001$

Now, let's imagine that *any* degree of accuracy $\varepsilon > 0$ that needs to be accommodated might be supplied ahead of time. Let's find such an n that a_n is within ε from $a = 0$. In other words, we need this inequality to be satisfied:

$$|a_n - a| = \left| \frac{1}{n^2} - 0 \right| = \frac{1}{n^2} < \varepsilon\,.$$

We solve it and choose N:

$$n > \frac{1}{\sqrt{\varepsilon}} = N\,.$$

This proves that the requirement can be satisfied. Then, for any such n we have $|a_n - a| < \varepsilon$, as required.

The problem is solved once and for all; the result gives us the same answers for the three particular choices of $\varepsilon = .1, .01, .001$ from the last example, as well as for any other! For example, if we pick $\varepsilon = .01$, what is N? By the formula, it is

$$N = \frac{1}{\sqrt{.01}} = \frac{1}{(10^{-2})^{1/2}} = \frac{1}{10^{-1}} = 10\,.$$

If we pick $\varepsilon = .0001$, what is N? By the formula, it is

$$N = \frac{1}{\sqrt{.0001}} = \frac{1}{(10^{-4})^{1/2}} = \frac{1}{10^{-2}} = 10^2 = 100\,.$$

Exercise 1.3.11

Carry out such an analysis for $a_n = 1/\sqrt{n}$.

Now the concept that most of calculus is built upon:

Definition 1.3.12: limit of sequence

We call number a the *limit* of a sequence a_n if the following condition holds:

▶ For each real number $\varepsilon > 0$, there exists a number N such that for every natural number $n > N$, we have:

$$|a_n - a| < \varepsilon\,.$$

We also say that *the limit is finite*. If a sequence has a limit, then we call the sequence *convergent* and say that it *converges*; otherwise it is *divergent* and we say it *diverges*.

Example 1.3.13: "for any ... there is"

The definition is logically complex; it is a combination of two familiar constructs. The first one:

- Two functions are equal, $f = g$, when `FOR ANY` x in the domain, we have $f(x) = g(x)$.
- `FOR ANY` number, its square cannot be negative.

- `FOR ANY` two odd numbers, their sum is even.

The second one:

- Two functions are unequal, $f \neq g$, when `THERE IS` such an x that $f(x) \neq g(x)$.
- `THERE ARE` no gaps in the graph of $y = x^2$.
- `THERE IS` a rectangular enclosure with the largest possible area when 100 yards of fencing is given.

This is how the two are often combined:

- `FOR ANY` number, `THERE IS` a larger number.
- `FOR ANY` two non-parallel lines, `THERE IS` an intersection.
- `THERE IS` a location on a rubber band that will remain in place `FOR ANY` combination of stretches and shrinks.

The statement in the definition of *limit* is even more complex with these phrases appearing three times:

▶ `FOR ANY` real number $\varepsilon > 0$, `THERE IS` a number N such that `FOR ANY` natural number $n > N$, we have
$$|a_n - a| < \varepsilon\,.$$

Exercise 1.3.14

Restate "`THERE IS` a rectangular enclosure with the largest possible area when 100 yards of fencing is given" as "`THERE IS` a rectangular enclosure so that `FOR ANY` ...".

Exercise 1.3.15

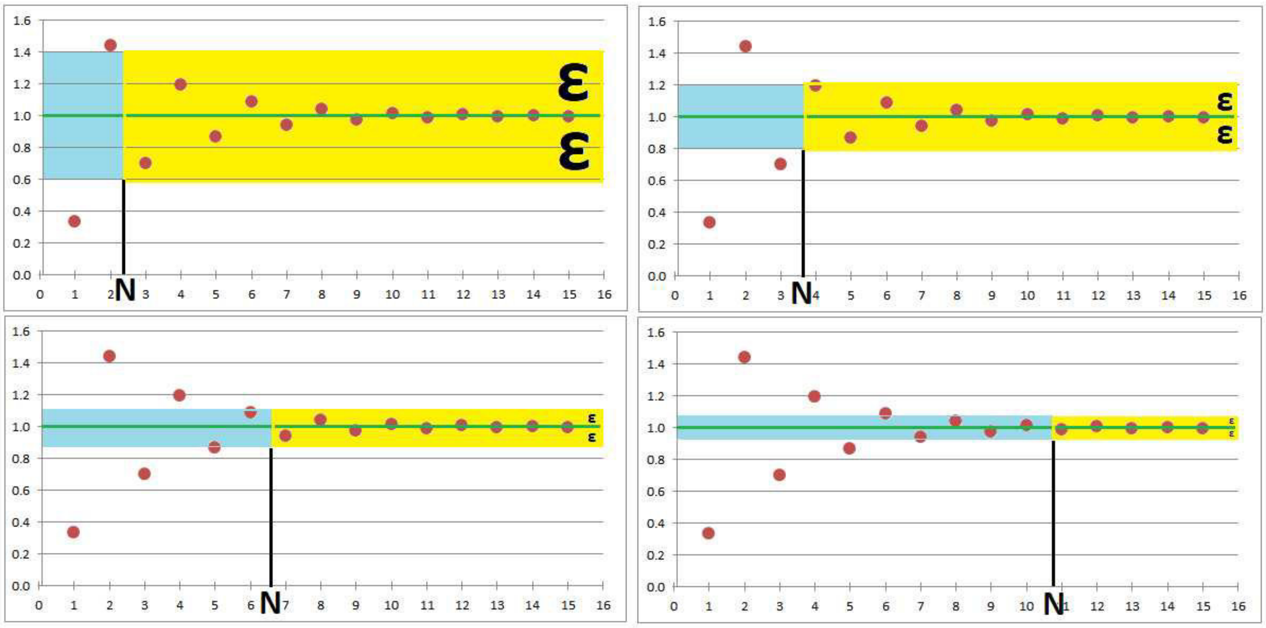
Suggest more instances of statements for the last example.

To prove the statement for a specific sequence, we will need to find such an N for each $\varepsilon > 0$.

Example 1.3.16: limit from definition

Let's apply the definition to
$$a_n = 1 + \frac{(-1)^n}{n}\,.$$
Suppose an $\varepsilon > 0$ is given. Looking at the numbers, we discover that they accumulate toward 1. Is this the limit? We apply the definition. Let's find such an n that a_n is within ε from $a = 1$:
$$|a_n - a| = \left|1 + \frac{(-1)^n}{n} - 1\right| = \left|\frac{(-1)^n}{n}\right| = \frac{1}{n} < \varepsilon\,.$$
We concentrate on the last part: For every ε, we need to point out such an N that for every $n > N$ we have:
$$\frac{1}{n} < \varepsilon\,.$$
Can we? Of course, the fraction will become smaller and smaller as we increase the denominator. Algebraically, we solve this inequality as follows:
$$n > \frac{1}{\varepsilon}\,.$$
That gives us the N required by the definition; we let
$$N = \frac{1}{\varepsilon}\,.$$
Then, for any $n > N$ we have $|a_n - a| < \varepsilon$, as required by the definition.

The way to visualize a trend in this convergent sequence is to enclose the end of the tail of the sequence in a *band*:

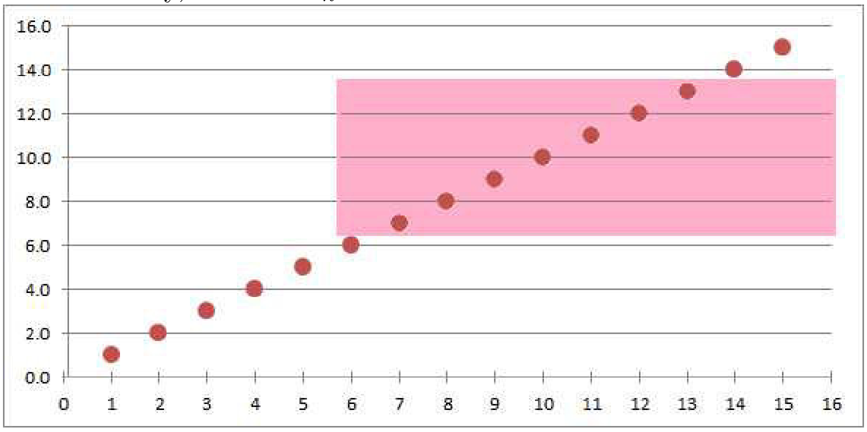


It should be, in fact, a narrower and narrower band; its width is 2ε. Meanwhile, the starting point of the band moves to the right; that's N.

Examples of divergence are below.

Example 1.3.17: divergence to infinity

A sequence may tend to infinity, such as $a_n = n$:



Then no band – no matter how wide – will contain the sequence's tail.

This behavior, however, has a meaningful pattern. But how do we explain that the sequence is approaching infinity? It will get over any threshold no matter how high:

Definition 1.3.18: infinite limit of sequence

We say that a sequence a_n *tends to positive infinity* if the following condition holds:

- ▶ For each real number R, there exists a natural number N such that for every natural number $n > N$, we have:

$$a_n > R\,.$$

We say that a sequence a_n *tends to negative infinity* if the following condition holds:

- ▶ For each real number R, there exists a natural number N such that for every natural number $n > N$, we have

$$a_n < R\,.$$

In either case, we also say that *the limit is infinite.*

Example 1.3.19: "for any ... there is"

The logic of the definition is the same as that of the last one:

▶ `FOR ANY` real number $R > 0$, `THERE IS` a number N such that `FOR ANY` natural number $n > N$, we have

$$a_n > R\,.$$

The following notation is routinely used:

Infinite limit of sequence

$$a_n \to \pm\infty$$

Example 1.3.20: divergence to infinity

Let's prove the last example $a_n = n$. We need to prove that:

▶ for each real number R, there exists a natural number N such that for every natural number $n > N$, we have:

$$a_n = n > R\,.$$

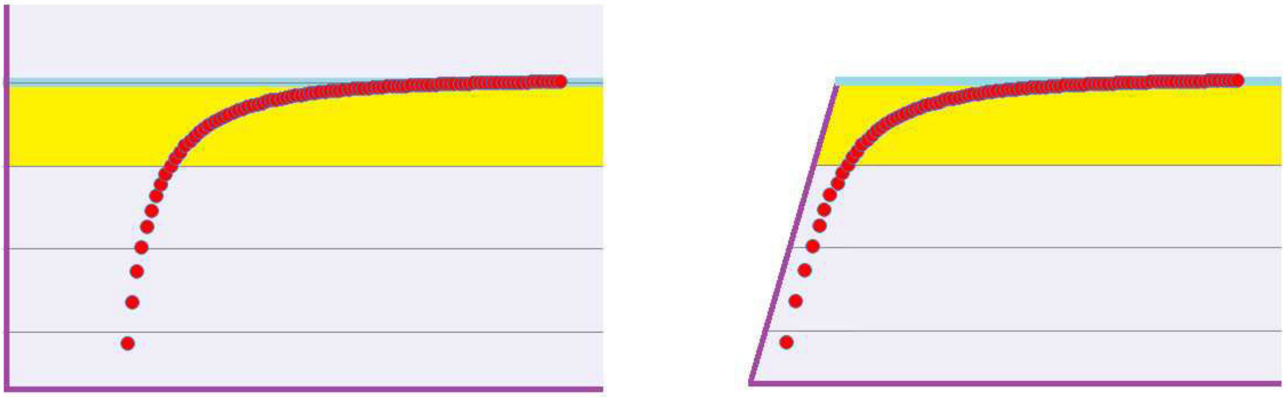
Easy, we choose $N = R$! Indeed:

$$n > N \implies a_n = n > N = R\,.$$

Exercise 1.3.21

Prove that the limits are infinite for the following sequences: (a) $-n$; (b) n^2; (c) $\sqrt{n}$.

Even though the two types of long-term behavior appear very different, there is a similarity in the geometry. Compare:



- On the left, the sequence approaches (but possibly never reaches) a particular line on the plane.
- On the right, the sequence approaches (but definitely never reaches) a certain imaginary line.

There are a total of three possibilities for every sequence!

Definition 1.3.22: convergence

Any sequence satisfies one of the three:

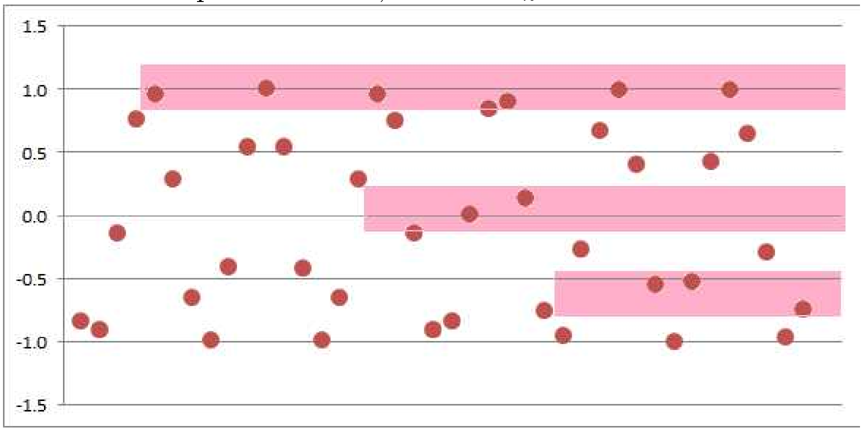
1. It has a finite limit, and we say it *converges.*
2. It has an infinite limit, and we say it *diverges to infinity* or simply that it

diverges.

3. It has neither a finite or infinite limit, and we say that *the limit does not exist* or simply that it *diverges*.

Example 1.3.23: no-pattern divergence

Some sequences seem to have no pattern at all, such as $a_n = \sin n$:



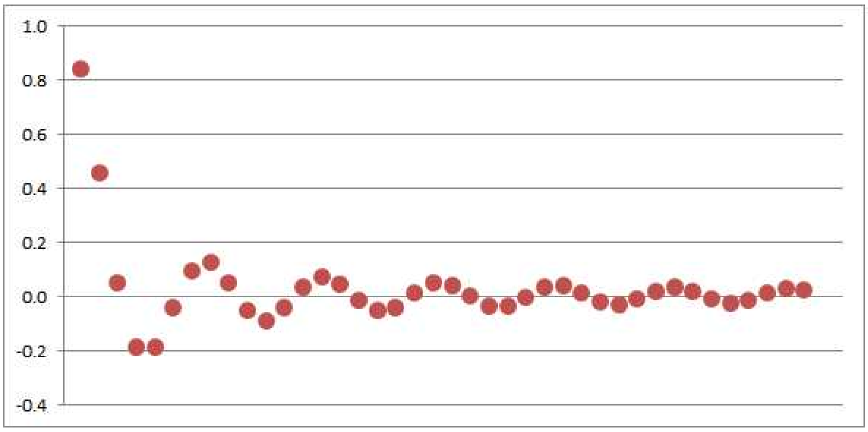
Here, no band – unless wide enough – can contain the sequence's tail. Nor does the sequence grow without bound.

Example 1.3.24: no-pattern convergence

If, however, we also divide this expression by n, resulting in

$$\frac{1}{n}\sin n\,,$$

the swings start to diminish:



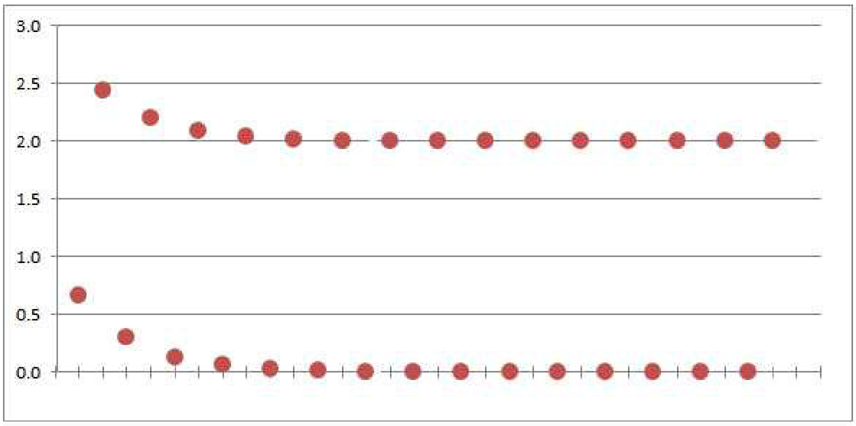
The limit is 0!

Example 1.3.25: two limits?

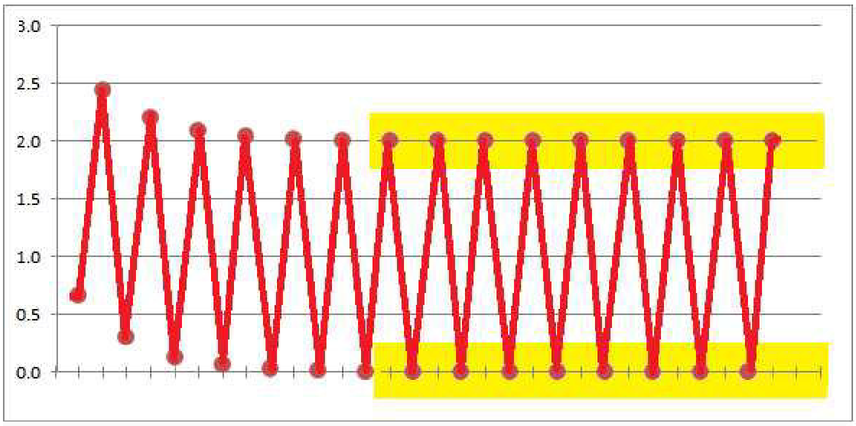
The next example is

$$a_n = 1 + (-1)^n + \frac{1}{n}\,.$$

It seems to approach *two limits* at the same time:



Indeed, no matter how narrow, we can find two bands to contain the sequence's two tails. But this means that no *single* band – if narrow enough – will contain them! The behavior turns out to be irregular:



Example 1.3.26: alternating sequence

Let's pick a simpler sequence and do this analytically. Let

$$a_n = (-1)^n = \begin{cases} 1 & \text{if } n \text{ is even,} \\ -1 & \text{if } n \text{ is odd.} \end{cases}$$

Is the limit $a = 1$? If it is, then this is what needs to be "small":

$$|a_n - a| = |(-1)^n - 1| = \begin{cases} 0 & \text{if } n \text{ is even,} \\ 2 & \text{if } n \text{ is odd.} \end{cases}$$

It's not! Indeed, this expression won't be less than ε if we choose it to be, say, 1, no matter what N is. So, $a = 1$ is not the limit. Is $a = -1$ the limit? Same story. In order to prove the negative, we need to try *every* possible value of a.

Exercise 1.3.27

Finish the proof in the last example.

Example 1.3.28: decimal approximations

For a given real number, we can construct a sequence that approximates that number – via *truncations* of its decimal approximations. For example, we have already seen this:

$$x_n = 0.9, 0.99, 0.999, 0.9999, \ldots \text{ tends to } 1\,.$$

Furthermore, we have:

$$x_n = 0.3, 0.33, 0.333, 0.3333, \ldots \text{ tends to } 1/3\,.$$

The idea of limit then helps us understand infinite decimals:

- What is the meaning of .9999...? It is the limit of the sequence $0.9, 0.99, 0.999, ...$; i.e., 1.
- What is the meaning of .3333...? It is the limit of the sequence $0.3, 0.3, 0.333, ...$; i.e., $1/3$.

Exercise 1.3.29

Find the nth-term formulas for the two sequences above and confirm the limits.

Example 1.3.30: definition of convergence

The definition of convergence is even more logically complex:

- `THERE IS` a number a such that
- `FOR ANY` real number $\varepsilon > 0$,
- `THERE IS` a number N such that
- `FOR ANY` natural number $n > N$,

we have:

$$|a_n - a| < \varepsilon\,.$$

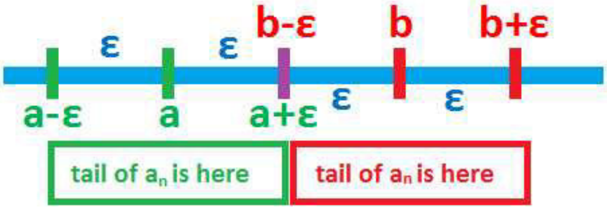
The definition speaks of *a number*! In every example, we found one number and then demonstrated that it satisfies the definition. But what if there are others? We need to justify “the” in “the limit”.

Theorem 1.3.31: Uniqueness of Limit of Sequence

A sequence can have only one limit (finite or infinite).

Proof.

The geometry of the proof is clear: We want to separate the two horizontal lines representing two potential limits by two non-overlapping bands, as shown above. Then the tail of the sequence would have to fit one or the other, but not both. These bands correspond to two intervals around those two “limits”. In order for them not to intersect, their width (that’s 2ε!) should be less than half the distance between the two numbers:



The proof is by contradiction. Suppose a and b are two limits, i.e., either satisfies the definition, and suppose also $a \neq b$. In fact, without loss of generality we can assume that $a < b$. Let

$$\varepsilon = \frac{b-a}{2}\,.$$

Then, what we are going to use at the end is

$$a + \varepsilon = b - \varepsilon\,.$$

Now, we rewrite the definition for a and b specifically:

▶ There exists a number L such that for every natural number $n > L$, we have

$$|a_n - a| < \varepsilon\,.$$

Now, we rewrite the definition for M as limit:

▶ There exists a number M such that for every natural number $n > M$, we have

$$|a_n - b| < \varepsilon\,.$$

In order to combine the two statements, we need them to be satisfied for the same values of n. Let

$$N = \min\{L, M\}\,.$$

We conclude the following:

- For every number $n > N$, we have
$$|a_n - a| < \varepsilon\,.$$
- For every number $n > N$, we have
$$|a_n - b| < \varepsilon\,.$$

In particular, for every $n > N$, we have:

$$a_n < a + \varepsilon = b - \varepsilon < a_n\,.$$

A contradiction.

Exercise 1.3.32

Follow the proof and demonstrate that it is impossible for a sequence to have as its limit: (a) a real number and $\pm\infty$, or (b) $-\infty$ and $+\infty$.

The theorem indicates that the correspondence makes sense:

$$\boxed{\text{a convergent sequence} \quad \longrightarrow \quad \text{its limit (a real number)}}$$

Can we reverse this correspondence? No, because there are many sequences converging to the same number:

$$\begin{array}{ccccc}
3+1/n & & 3-1/n & & 3+2/n \\
 & \searrow & \downarrow & \swarrow & \\
3+1/n^2 & \to & \boxed{3} & \leftarrow & 3+(-1)^n/n \\
 & \nearrow & \uparrow & \nwarrow & \\
3+1/2^n & & 3-1/\sqrt{n} & & 3+1/\ln n
\end{array}$$

However, we can say that a real number "is" its approximations, i.e., all sequences that converge to it.

Thus, there can be no two limits of a sequence and we are justified to speak of *the* limit. If it exists, it's either a number or one of the two infinities.

We use the following notation for the limit of a sequence:

Limit

$$a_n \to a \ \text{ as } n \to \infty$$

and

$$\lim_{n\to\infty} a_n = a$$

It reads "the limit of a_n is a" or "a_n converges to a".

The theorem can then be restated as follows (we will use $\Longrightarrow$ for an implication, i.e., an "if-then" statement):

Corollary 1.3.33: Uniqueness of Limit of Sequence

If we have two limits of the same sequence, they are equal.

In other words, we have:

$$\boxed{a = \lim_{n\to\infty} x_n \quad \texttt{AND} \quad b = \lim_{n\to\infty} x_n \implies a = b}$$

Exercise 1.3.34

State as an implication: (a) "When there is smoke, there is fire." (b) "Every square is a rectangle."

We use a similar notation to describe the particular kind of divergent behavior (infinity):

Infinite limit

$$a_n \to \pm\infty \quad \text{as } n \to \infty$$

or

$$\lim_{n\to\infty} a_n = \pm\infty$$

It reads "the limit of a_n is infinite" or "a_n diverges to infinity".

The third possibility is that there is no trend. Since the definition of convergence starts with "`THERE IS` a number a...", we can say that otherwise there is no limit:

No limit

$$\lim_{n\to\infty} a_n, \quad \text{no limit}$$

or

$$\lim_{n\to\infty} a_n \quad \text{DNE}$$

It reads "the limit of a_n does not exist".

Example 1.3.35: notation

We can now write the conclusions in the above examples using our new notation, as follows:

$$\begin{array}{llllllll}
1 & 1/2 & 1/3 & 1/4 & 1/5 & \dots & \to 0\,. \\
.9 & .99 & .999 & .9999 & .99999 & \dots & \to 1\,. \\
1. & 1.1 & 1.01 & 1.001 & 1.0001 & \dots & \to 1\,. \\
3. & 3.1 & 3.14 & 3.141 & 3.1415 & \dots & \to \pi\,. \\
1 & 2 & 3 & 4 & 5 & \dots & \to +\infty\,. \\
0 & 1 & 0 & 1 & 0 & \dots & \text{no limit}.
\end{array}$$

The limits of some specific sequences (seen in Volume 1, Chapter 1PC-1) can be easily found as shown in the two formulas below.

Theorem 1.3.36: Limit of Constant Sequence

For any real c, we have:

$$\boxed{\lim_{n\to\infty} c = c}$$

Theorem 1.3.37: Limit of Power Sequence

For any integer k, we have:

$$\boxed{\lim_{n\to\infty} n^k = \begin{cases} 0 & \text{if } k<0 \\ 1 & \text{if } k=0 \\ +\infty & \text{if } k>0 \end{cases}}$$

Proof.

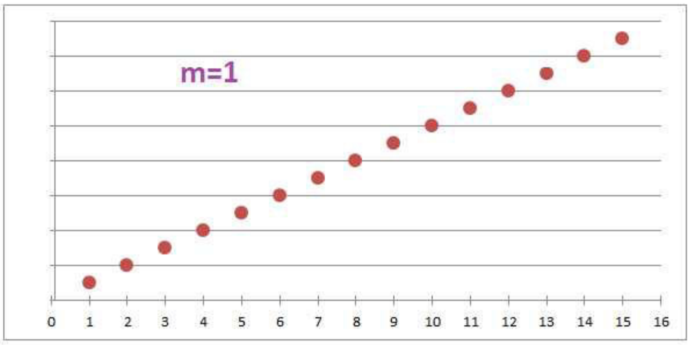
First, the case of $k<0$. Suppose $\varepsilon>0$ is given. We need to find such an N that $|n^k-0| = n^k < \varepsilon$ whenever $n>N$. We can express such an N in terms of this ε; we just choose:

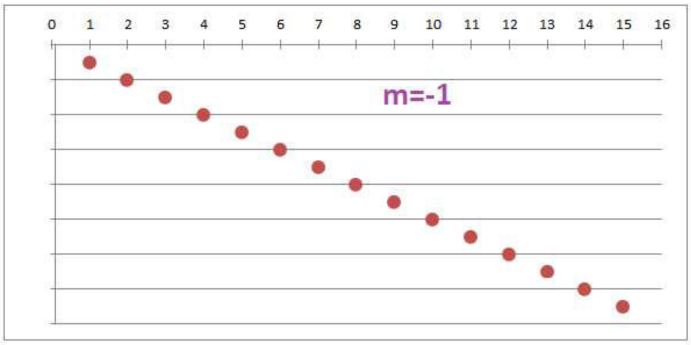
$$N = \sqrt[1/k]{\varepsilon}\,.$$

Second, the case of $k>0$. Suppose $R>0$ is given. We need to find such an N that $n^k > R$ whenever $n>N$. We can express such an N in terms of this R; similarly, to the above we choose:

$$N = \sqrt[1/k]{R}\,.$$

This is what a few typical arithmetic progressions look like:





The theorem below summarizes these observations:

Theorem 1.3.38: Limit of Arithmetic Progression

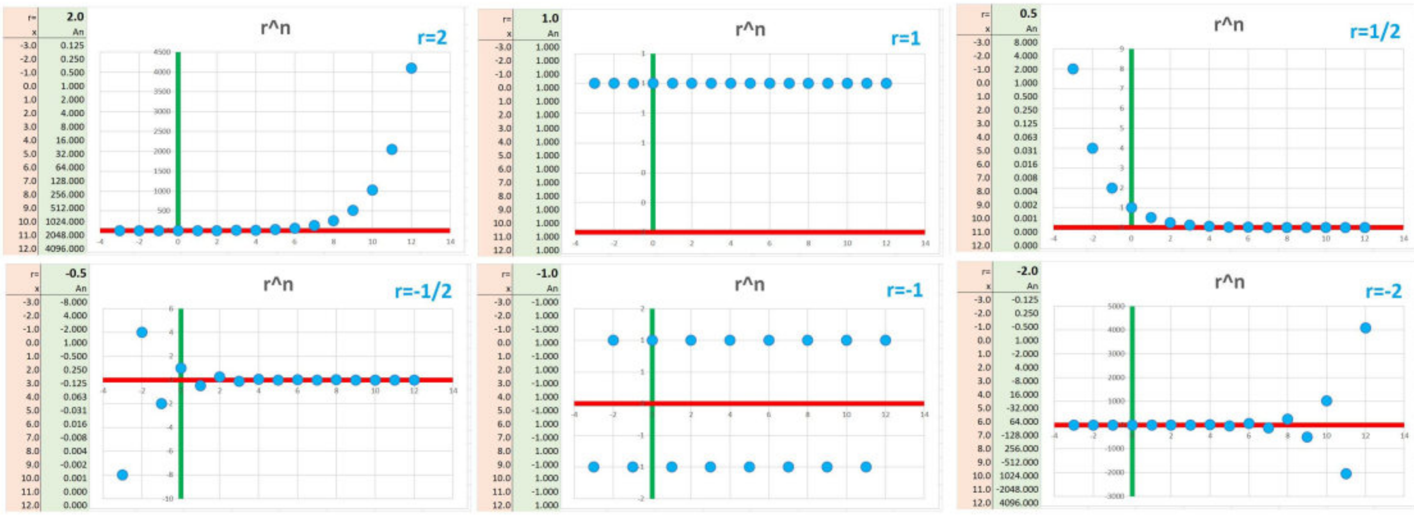
For any real numbers $m, b>0$, we have:

$$\boxed{\lim_{n\to\infty} (b+nm) = \begin{cases} -\infty & \text{if } m<0 \\ b & \text{if } m=0 \\ +\infty & \text{if } m>0 \end{cases}}$$

Exercise 1.3.39

Prove the theorem.

This is what a few typical geometric progressions look like:



The theorem below summarizes these observations:

Theorem 1.3.40: Limit of Geometric Progression

For any real number r, we have:

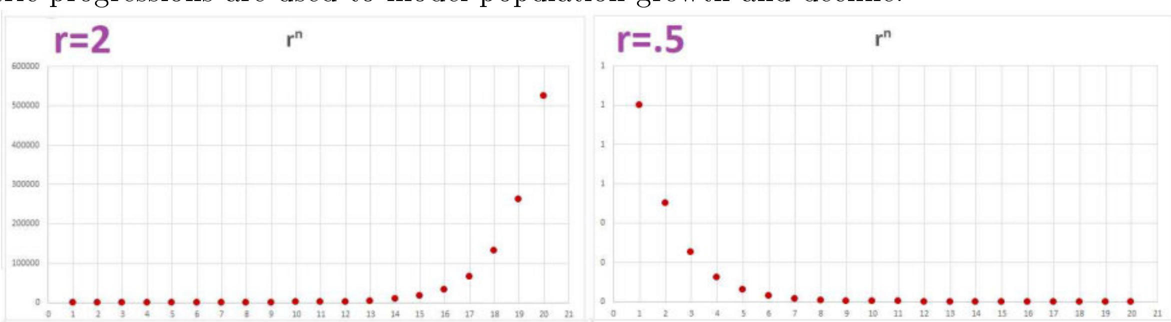
$$\boxed{\lim_{n\to\infty} r^n = \begin{cases} \textit{no limit} & \textit{if } r \le -1 \\ 0 & \textit{if } |r| < 1 \\ 1 & \textit{if } r = 1 \\ +\infty & \textit{if } r > 1 \end{cases}}$$

Exercise 1.3.41

Prove the theorem.

Example 1.3.42: population growth and decline

Geometric progressions are used to model population growth and decline:



We conclude from the theorem that:

- In case of growth, the population will become as large as we desire if we wait long enough.
- In case of decline, the population will become as small as we desire if we wait long enough.

The fact that in the latter case we will have a population of less than one person but never zero indicates a limitation of this model.

Exercise 1.3.43

Find the limit of each of these sequences or show that it doesn't exist:

1. $1,\ 3,\ 5,\ 7,\ 9,\ 11,\ 13,\ 15,\ \ldots$
2. $.9,\ .99,\ .999,\ .9999,\ \ldots$
3. $1,\ -1,\ 1,\ -1,\ \ldots$
4. $1,\ 1/2,\ 1/3,\ 1/4,\ \ldots$

5. $1,\ 1/2,\ 1/4\,, 1/8,\ \ldots$
6. $2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ \ldots$
7. $1,\ -4,\ 9,\ -16,\ 25,\ \ldots$
8. $3,\ 1,\ 4,\ 1,\ 5,\ 9,\ \ldots$

Example 1.3.44: series

In either of the two tables below, we have a sequence given in the first two columns. Its nth-term formula is known. The third column shows the sequence of sums (seen in Volume 1, Chapter 1PC-1) of the first:

n	a_n	s_n
1	$\frac{1}{1}$	$\frac{1}{1}$
2	$\frac{1}{2}$	$\frac{1}{1}+\frac{1}{2}$
3	$\frac{1}{3}$	$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}$
$\vdots$	$\vdots$	$\vdots$
n	$\frac{1}{n}$	$\frac{1}{1}+\frac{1}{2}+\frac{1}{3}+\ldots+\frac{1}{n}$

n	a_n	s_n
1	$\frac{1}{1}$	$\frac{1}{1}$
2	$\frac{1}{2}$	$\frac{1}{1}+\frac{1}{2}$
3	$\frac{1}{4}$	$\frac{1}{1}+\frac{1}{2}+\frac{1}{4}$
$\vdots$	$\vdots$	$\vdots$
n	$\frac{1}{2^n}$	$\frac{1}{1}+\frac{1}{2}+\frac{1}{4}+\ldots+\frac{1}{2^{n-1}}$

The nth-term formula is unknown as we don't know how to represent these quantities without "...". In contrast to the last example, finding the limit of such a sequence is a challenge (Volume 3, Chapter 3IC-5).

Exercise 1.3.45

Give formulas for the following sequences: (a) $a_n \to 0$ as $n \to \infty$ but it's not increasing or decreasing; (b) $b_n \to +\infty$ as $n \to \infty$ but it's not increasing.

1.4. Limits under algebraic operations

If every real number *is* the sequence of its approximations, does the usual arithmetic operations with numbers match those with the corresponding sequences? Yes. We will discover the following:

▶ Limits behave well with respect to algebra.

For simplicity, we assume below that all the sequences are defined on the same set of integers.

Example 1.4.1: algebra of sequences

What do we mean by adding, multiplying, etc. two sequences? Just as with functions, we add, multiply, etc. term-wise:

n	1	2	3	...	n	...
a_n	1	$1/2$	$1/3$	...	$1/n$	...
b_n	-1	1	-1	...	$(-1)^n$	...
a_n+b_n	$1-1$	$1/2+1$	$1/3-1$	...	$1/n+(-1)^n$	...
$a_n\cdot b_n$	$1\cdot 1$	$1/2\cdot 1$	$1/3\cdot(-1)$	...	$1/n\cdot(-1)^n$	...

Of course, we really need only the nth column.

The three main classes of sequences – power, arithmetic, and geometric – introduced in the last section will serve as “seeds”. When combined via algebraic operations, they produce an infinite variety of sequences, the limits of which we should be able to find.

We will start with our study of convergence of sequences in general with a special class of simpler sequences.

Below we have a sequence a_n plotted, as well as the sequence of the distances from a_n to a, i.e., $b_n = |a_n - a|$:

The theorem below shows why this is important (we will use “$\Longleftrightarrow$” to indicate an equivalence, i.e., an “if-and-only-if” statement):

Theorem 1.4.2: Convergence to Zero

A sequence converges to a number if and only if the limit of the distance from the terms of this sequence to this number is zero.

In other words, for any sequence a_n, we have:

$$\boxed{a_n \to a \iff |a_n - a| \to 0}$$

Using the other notation for limits, we have:

$$\boxed{\lim_{n\to\infty} a_n = a \iff \lim_{n\to\infty} |a_n - a| = 0}$$

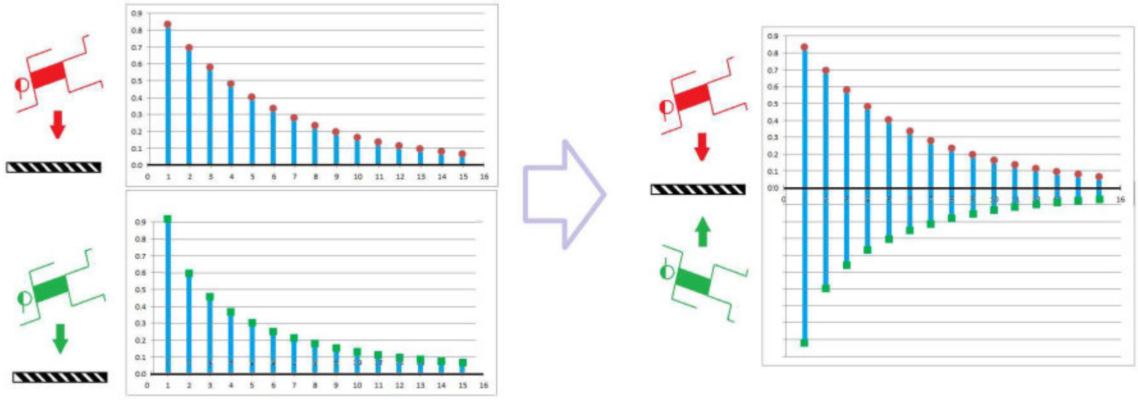
So, if the limit of a_n is a, then the limit of $|a_n - a|$ is 0, and vice versa.

Then, to understand limits of sequences in general, we need first to understand those of this smaller class: *positive sequences that converge to* 0. The definition of convergence becomes simpler! Indeed, $0 < a_n \to 0$ when

- for any $\varepsilon > 0$, there is an N such that $a_n < \varepsilon$ for all $n > N$.

As we know (Volume 1, Chapter 1PC-3), some simple algebraic operations on the outputs of function produce transformations of the plane and vice versa. But how do they affect the limits? Predictably.

We start with addition. If two people are walking away from a pole in the opposite directions, we have two sequences a_n and b_n that represent their locations (left). What is the sum of the two sequences $a_n + b_n$ (right)? The distance between them:



To graphically add two sequences, we flip the second upside down and then connect each pair of dots with a bar. Then, the lengths of these bars form the new sequence. Now, if the original bars diminish, then so do the new ones, as follows:

Theorem 1.4.3: Sum Rule for Zero Limit Sequences

If either of two sequences with non-negative terms converges to zero, then so does their sum.

In other words, we have:

$$0 \le a_n \to 0 \quad \texttt{AND} \quad 0 \le b_n \to 0 \implies a_n + b_n \to 0\,.$$

Proof.

Suppose $\varepsilon > 0$ is given. From the definition,

- $a_n \to 0 \implies$ there is N such that $a_n < \varepsilon/2$ for all $n > N$, and
- $b_n \to 0 \implies$ there is M such that $b_n < \varepsilon/2$ for all $n > M$.

Then for all $n > \max\{N, M\}$, we have:

$$a_n + b_n < \varepsilon/2 + \varepsilon/2 = \varepsilon\,.$$

Therefore, by definition, we have: $a_n + b_n \to 0$.

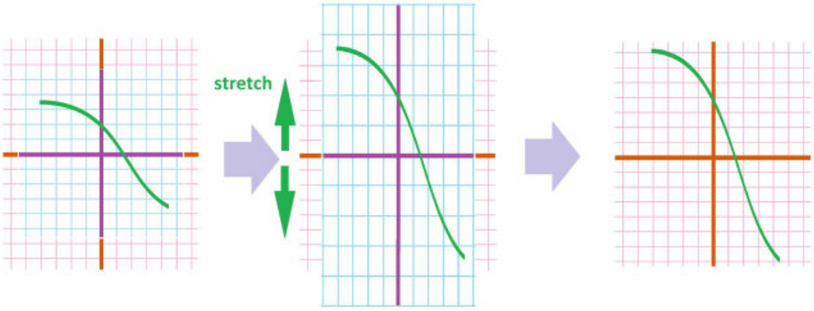
Exercise 1.4.4

Prove the version of the above theorem for m sequences (a) from the theorem and (b) by generalizing the proof.

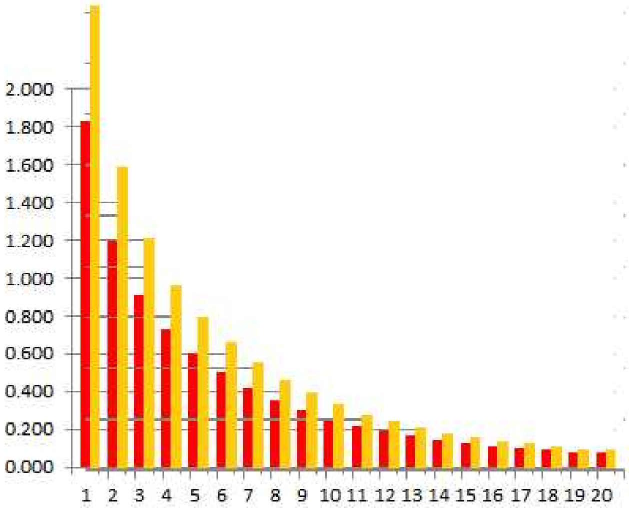
Example 1.4.5: application of Sum Rule

$$\frac{1}{n} \to 0 \quad \texttt{AND} \quad \frac{1}{n^2} \to 0 \implies \frac{1}{n} + \frac{1}{n^2} \to 0\,.$$

Before considering the product of two sequences, let's consider what happens when we multiply a sequence by a constant number. We know that such multiplication simply stretches the whole plane in the vertical direction (both up and down, away from the n-axis).



The graph also stretches. Here we see a_n (red) and $1.2 \cdot a_n$ (orange):



Under such stretch, zero remains zero!

Theorem 1.4.6: Constant Multiple Rule for Zero Limit Sequences

If a sequence with non-negative terms converges to zero, then so does any of its multiples.

In other words, we have:

$$0 \le a_n \to 0 \implies ca_n \to 0 \text{ for any real } c > 0\,.$$

Proof.

Suppose $\varepsilon > 0$ is given. From the definition, we have:

- $0 < a_n \to 0 \implies$ there is N such that $a_n < \varepsilon/c$ for all $n > N$.

Then for all $n > N$, we have:

$$c \cdot a_n < c \cdot \varepsilon/c = \varepsilon\,.$$

Therefore, by definition, we have: $ca_n \to 0$.

Example 1.4.7: application of Constant Multiple Rule

$$\frac{1}{n} \to 0 \implies \frac{5}{n} \to 0\,.$$

For more complex situations, we need to use the fact that convergent sequences are *bounded* as functions (seen in Volume 1, Chapter 1PC-2):

Theorem 1.4.8: Boundedness of Convergent Sequences

A convergent sequence is bounded.

In other words, we have:

$$a_n \to a \implies |a_n| < Q \text{ for some real } Q\,.$$

Proof.

The idea is that the tail of the sequence fits into a (narrow) band; meanwhile, there are only finitely many terms left! Choose $\varepsilon = 1$. Then by definition, there is such N that for all $n > N$ we have:

$$|a_n - a| < 1\,.$$

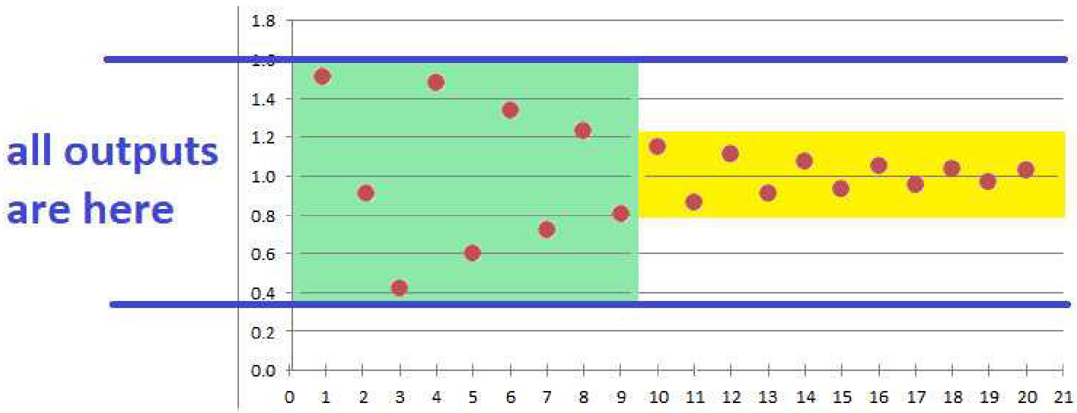
Then, we have:

$$\begin{array}{lll} |a_n| & = |(a_n - a) + a| & \text{We add and subtract } a. \\ & \leq |a_n - a| + |a| & \text{We use the Triangle Inequality (Chapter 1PC-5).} \\ & < 1 + |a|. & \text{We use the inequality above.} \end{array}$$

To finish the proof, we choose:

$$Q = \max\left\{|a_1|, ..., |a_N|, 1 + |a|\right\}.$$

So, the sequence fits into a (not necessarily narrow) band. The proof is illustrated below:



The converse isn't true: Not every bounded sequence is convergent. Just try $a_n = \sin n$. We will show later that, with an extra condition, bounded sequences do have to converge.

A smaller sequence would be squeezed between the larger one and zero:

Corollary 1.4.9: Squeeze Theorem for Zero Limit Sequences

If a sequence with non-negative terms converges to zero, then so does any other smaller sequence with non-negative terms.

In other words, we have:

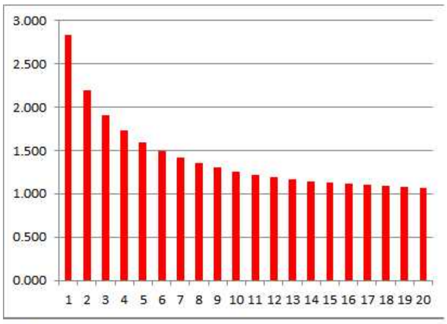
$$0 < a_n \to 0 \ \texttt{AND} \ \ 0 < b_n < a_n \implies b_n \to 0\,.$$

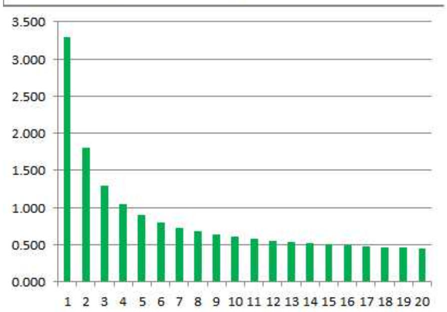
Exercise 1.4.10

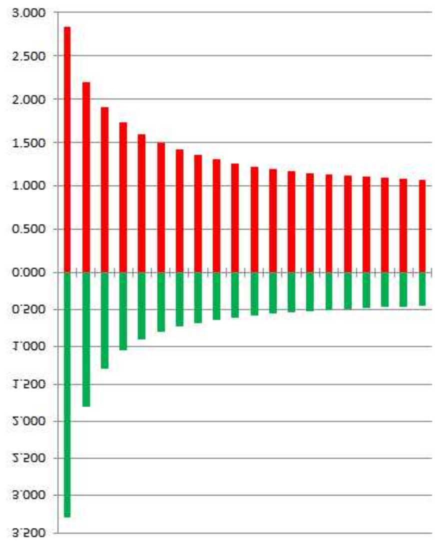
Prove the corollary.

We are now ready for the *general* results on the algebra of limits.

First, the summation:







Theorem 1.4.11: Sum Rule for Limits of Sequences

If sequences a_n, b_n converge, then so does $a_n + b_n$.

Furthermore, we have:

$$\boxed{\lim_{n\to\infty} (a_n + b_n) = \lim_{n\to\infty} a_n + \lim_{n\to\infty} b_n}$$

Proof.

Suppose

$$a_n \to a,\ b_n \to b\,.$$

Then

$$|a_n - a| \to 0,\ |b_n - b| \to 0\,.$$

We compute

$$\begin{array}{lll} |(a_n + b_n) - (a + b)| & = |(a_n - a) + (b_n - b)| & \text{We rearrange the terms.} \\ & \leq |a_n - a| + |b_n - b| & \text{We use the Triangle Inequality.} \\ & \to 0 + 0 & \text{We use Sum Rule for Zero Limit Sequences.} \\ & = 0\,. & \end{array}$$

Then, by the last corollary, we have:

$$|(a_n + b_n) - (a + b)| \to 0\,.$$

Then, by the first theorem, we have:

$$a_n + b_n \to a + b\,.$$

Warning!

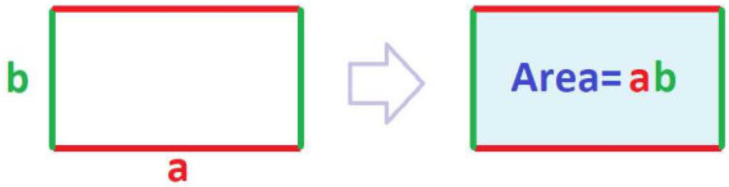
There is one condition (convergence) and two conclusions (convergence and the value of the limit of the sum).

Exercise 1.4.12

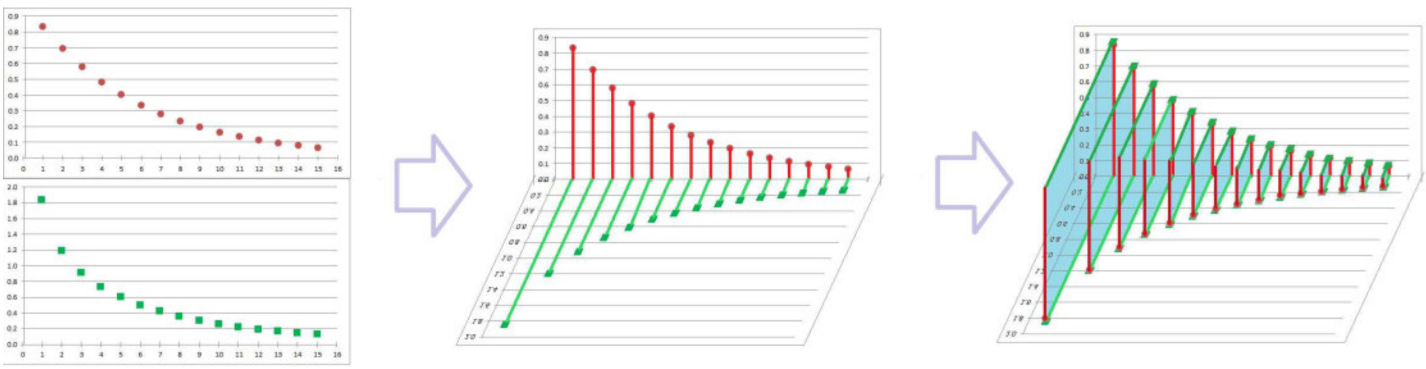
What if one or both of the limits are infinite?

As you can see, such a limit is simply “split” in half.

While the geometric interpretation of addition is putting two bars together, what is that for multiplication? If a and b are the sides of a rectangle, then ab is its area.



So, when two sequences are multiplied, it is as if we use each pair of their values to build a rectangle:



Then the areas of these rectangles form a new sequence and these areas converge if the widths and the heights converge:

Theorem 1.4.13: Product Rule for Limits of Sequences

If sequences a_n, b_n converge, then so does $a_n \cdot b_n$.

Furthermore, we have:

$$\boxed{\lim_{n\to\infty}(a_n \cdot b_n) = \Big(\lim_{n\to\infty} a_n\Big) \cdot \Big(\lim_{n\to\infty} b_n\Big)}$$

Proof.

Suppose

$$a_n \to a,\ b_n \to b.$$

Then,

$$|a_n - a| \to 0,\ |b_n - b| \to 0\,.$$

We compute:

$$\begin{array}{lll}
|a_n \cdot b_n - a \cdot b| & = |a_n \cdot b_n + (-a \cdot b_n + a \cdot b_n) - a \cdot b| & \text{We add extra terms and then factor.}\\
& = |(a_n - a) \cdot b_n + a \cdot (b_n - b)| & \\
& \le |(a_n - a) \cdot b_n| + |a \cdot (b_n - b)| & \text{We use the Triangle Inequality.}\\
& = |a_n - a| \cdot |b_n| + |a| \cdot |b_n - b| & \\
& \le |a_n - a| \cdot Q + |a| \cdot |b_n - b| & \text{We use the Boundedness Theorem.}\\
& \to 0 \cdot Q + |a| \cdot 0 & \text{We use SR and CMR.}\\
& = 0. &
\end{array}$$

Therefore,

$$a_n \cdot b_n \to a \cdot b\,.$$

The limit splits in half, again.

Exercise 1.4.14

What if one or both of the limits are infinite?

The Constant Multiple Rule follows from the theorem:

Corollary 1.4.15: Constant Multiple Rule for Limits of Sequences

If sequence a_n converges, then so does ca_n for any real c.

Furthermore, we have:

$$\boxed{\lim_{n\to\infty}\left(c\cdot a_n\right)=c\cdot\lim_{n\to\infty}a_n}$$

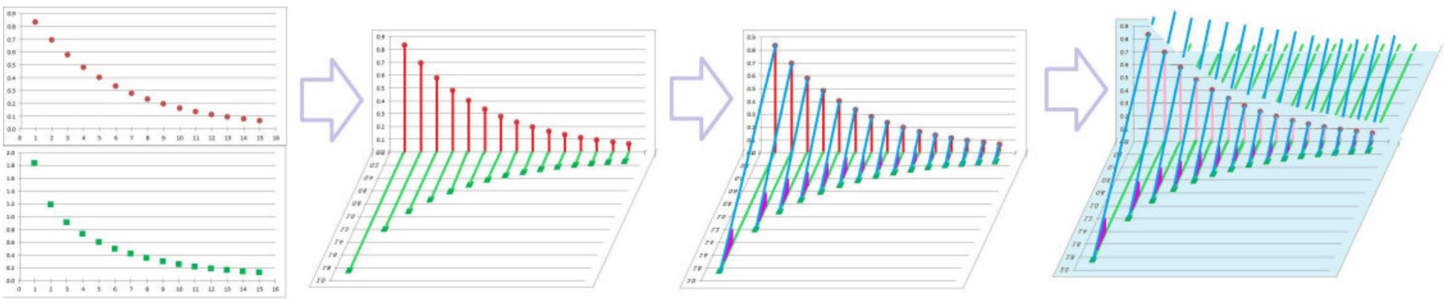
As you can see, a constant is simply "factored" out of a (convergent!) limit.

Exercise 1.4.16

Derive the corollary from the last theorem.

One can understand division of sequences as multiplication in reverse: If the areas of the rectangles converge and so do their widths, then so do their heights.

Also, when two sequences are divided, it is as if we use each pair of their values to build a triangle:



Then the tangents of the base angles of these triangles form a new sequence and they converge if the widths and the heights converge. This result is especially important in the forthcoming parts of calculus:

Theorem 1.4.17: Quotient Rule for Limits of Sequences

If sequences a_n, b_n converge, then so does a_n/b_n whenever defined, and

$$\boxed{\lim_{n\to\infty}\left(\frac{a_n}{b_n}\right)=\frac{\displaystyle\lim_{n\to\infty}a_n}{\displaystyle\lim_{n\to\infty}b_n}}$$

provided $\lim\limits_{n\to\infty}b_n\neq 0\,.$

Proof.

We will only prove the case of $a_n=1$. Suppose $b_n\to b\neq 0$. First, choose $\varepsilon=|b|/2$ in the definition of convergence. Then there is N such that for all $n>N$ we have:

$$|b_n-b|<|b|/2\,.$$

Therefore,

$$|b_n| > |b|/2\,.$$

Next,

$$\begin{aligned}
\left|\frac{1}{b_n} - \frac{1}{b}\right| &= \left|\frac{b-b_n}{b_n b}\right| \\
&= \frac{|b-b_n|}{|b_n|\cdot|b|} \\
&< \frac{|b-b_n|}{|b/2|\cdot|b|} && \text{From the above inequality.} \\
&\to \frac{0}{|b/2|\cdot|b|} && \text{We use CMR.} \\
&= 0\,.
\end{aligned}$$

Therefore,

$$\frac{1}{b_n} \to \frac{1}{b}\,.$$

Finally, the general case of the *Quotient Rule* follows from *Product Rule*:

$$\frac{a_n}{b_n} = a_n \cdot \frac{1}{b_n} \to a \cdot \frac{1}{b} = \frac{a}{b}\,.$$

Warning!

There are two conditions (convergence and non-zero of the limit of the denominator) and two conclusions (convergence and the value of the limit of the ratio).

Exercise 1.4.18

What if one or both of the limits are infinite?

The summary result below shows that when we replace every real number with a sequence converging to it, it is still possible to do algebraic operations with them:

Theorem 1.4.19: Algebra of Limits of Sequences

Suppose $a_n \to a$ *and* $b_n \to b$. *Then*

SR:	$a_n + b_n \to a + b$	*CMR:*	$c \cdot a_n \to ca$	*for any real* c
PR:	$a_n \cdot b_n \to ab$	*QR:*	$a_n/b_n \to a/b$	*provided* $b \neq 0$

As you can see, (convergent!) sequences are simply replaced with their limits. These results are also known as the *Limits Theorems*.

Example 1.4.20: using limits rules

Let

$$a_n = 7n^{-2} + \frac{2}{3^n} + 8\,.$$

What is its limit as $n \to \infty$? We recognize three "seed" sequences: n^{-2} is a power sequence, 3^n is geometric, and 8 is constant. These sequences with known limits, however, are combined by algebra.

The computation of the limit below is straightforward, but every step has to be justified with the rules of limits presented above. To understand which rule to apply *first*, observe what the *last* operation is. It is addition. We, therefore, use the *Sum Rule* first, subject to justification:

$$\begin{aligned}
\lim_{n\to\infty} a_n &= \lim_{n\to\infty}(7n^{-2}+\frac{2}{3^n}+8) && \text{We use SR.}\\
&= \lim_{n\to\infty}(7\cdot n^{-2})+\lim_{n\to\infty}(2\cdot\frac{1}{3^n})+\lim_{n\to\infty}8 && \text{We use CMR.}\\
&= 7\cdot\lim_{n\to\infty}n^{-2}+2\cdot\lim_{n\to\infty}3^{-n}+8 && \text{We use the limits from the last section.}\\
&= 7\cdot 0+2\cdot 0+8\\
&= 8\,.
\end{aligned}$$

As all the limits exist, our use of the *Sum Rule* (and then the *Constant Multiple Rule*) was justified.

Warning!

It is considered a serious error if you use the conclusion (the formula) of one of these rules without verifying the conditions (the convergence of the sequences involved).

Example 1.4.21: infinite limits

Prove the limit:

$$\lim_{n\to\infty}(n^2-n)=+\infty\,.$$

Plotting the graph does suggest that the limit is infinite.

Since the *last* operation is addition, we are supposed to use the *Sum Rule* first. However, *both* of the terms go to ∞, which makes the *Sum Rule* inapplicable. An attempt to apply the *Sum Rule* – over this objection – would result in a *meaningless* expression:

$$n^2-n \xrightarrow{???} \infty-\infty\,.$$

We can't conclude that the limit doesn't exist, or that is does! We've just failed to find the answer. The issue is resolved in the next section.

Presented verbally, these rules have these abbreviated versions:

- The limit of the *sum* is the sum of the limits.
- The limit of the *difference* is the difference of the limits.
- The limit of the *product* is the product of the limits.
- The limit of the *quotient* is the quotient of the limits (as long as that of the denominator isn't zero).

We can also represent these rules as diagrams. For example, this is the Sum Rule:

$$\begin{tikzcd}
a_n, b_n \arrow[r, "\lim"] \arrow[d, "+"] \arrow[dr, phantom, "SR"] & a, b \arrow[d, "+"] \\
a_n+b_n \arrow[r, "\lim"] & \lim(a_n+b_n)=a+b
\end{tikzcd}$$

In the diagram, we start with a pair of sequences at the top left and then we proceed in two ways:

- right: take the limit of either, then down: add the results; or
- down: add them, then right: take the limit of the result.

The result is the same! For the *Product Rule* and the *Quotient Rule*, we just replace "+" with "·" and "÷" respectively.

The following combination of the Sum Rule and the Constant Multiple Rule is worth remembering:

Theorem 1.4.22: Linearity Rule for Limits of Sequences

Suppose $a_n \to a$ and m and b are real numbers. Then

$$\boxed{ma_n + b \to ma + b}$$

What about *infinite limits*? If we replace infinity (positive or negative) with a sequence that approaches it, will the algebra make sense?

1.5. Can we add infinities? Subtract? Divide? Multiply?

We have demonstrated that in our computations of limits we can replace any sequence with its limit and continue doing the algebra. This conclusion doesn't apply to divergent sequences!

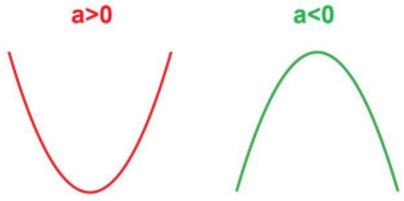
Warning!

Never forget to confirm the preconditions when using these rules.

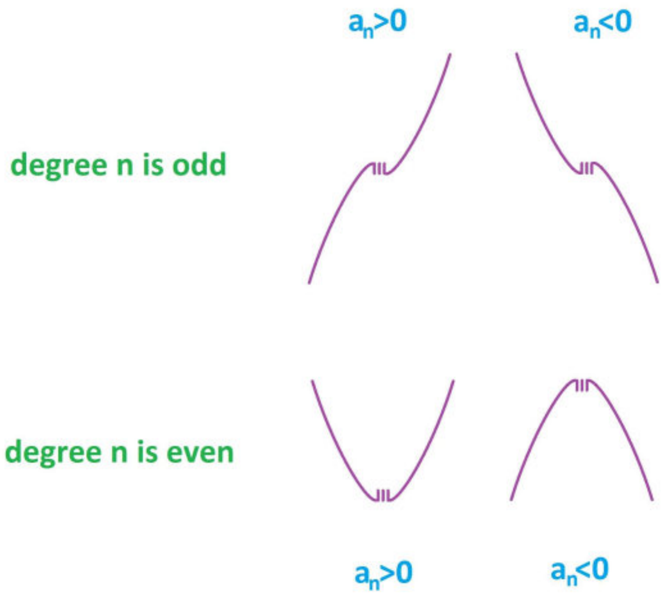
Sequences that approach infinity *diverge*, technically, but they provide useful information about the pattern exhibited by the sequences. Such a sequence can also be used to create a convergent sequence:

$$a_n = m \ \text{ and } \ b_n = \frac{1}{n}\,.$$

Recall that the direction of the graph of a quadratic polynomial $f(x) = ax^2 + bx + c$ (a parabola) is determined by the sign of a only:



The picture suggests that the limits are infinite, $+\infty$ in the former case and $-\infty$ in the latter. We even know more: The leading term will determine the large scale shape of the graph of any polynomial $a_p n^p + a_{p-1}n^{p-1} + ... + a_1 n + a_0$:



The following result makes these ideas precise:

Theorem 1.5.1: Limits of Polynomials at Infinity

Suppose we have a polynomial of degree p (the leading coefficient $a_p \neq 0$). Then the limit of the sequence defined by this function is:

$$\lim_{n\to\infty}\left(a_p n^p + a_{p-1}n^{p-1} + ... + a_1 n + a_0\right) = \begin{cases} +\infty & \text{if } a_p > 0 \\ -\infty & \text{if } a_p < 0 \end{cases}$$

Proof.

The idea is to *factor out the highest power*:

$$a_p n^p + a_{p-1}n^{p-1} + ... + a_1 n + a_0 = n^p \cdot \left(a_p + a_{p-1}n^{-1} + ... + a_1 n^{1-p} + a_0 n^{-p}\right).$$

Then the limit of the first factor is ∞ and that of the second is $a_p + 0 = a_p$.

So, as far as its behavior at ∞, for a polynomial, *only the leading term matters*:

$$\lim_{n\to\infty}\left(\boxed{\mathbf{a_p n^p}} + a_{p-1}n^{p-1} + ... + a_1 n + a_0\right) = \lim_{n\to\infty}\boxed{\mathbf{a_p n^p}} = \begin{cases} +\infty & \text{if } a_p > 0 \\ -\infty & \text{if } a_p < 0 \end{cases}$$

From polynomials to rational functions:

Definition 1.5.2: rational function

The ratio of two polynomials is called a *rational function*.

Warning!

All polynomials are rational functions too.

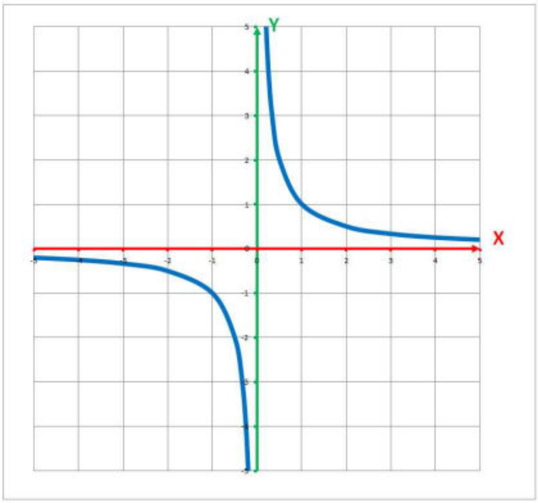
The problems we face are different.

Example 1.5.3: reciprocal sequence

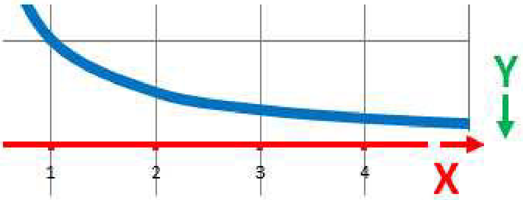
We are already familiar with such an important sequence as the reciprocals:

$$a_n = \frac{1}{n}\,.$$

Even with such a simple formula, once division is introduced, the complexity increases dramatically. We can see – in comparison to the polynomial sequences – some new features in the graph ($y = 1/x$):



Let's look at the x-axis. If the graph can't cross it, it has to start to *crawl*, with virtually no up or down progress:



The phenomenon is also seen in the data:

n	1	2	3	...	100	...
$1/n$	1	$1/2$	$1/3$	...	$1/100$	...

If we zoom out, the ends of the graph merge with the axis. In other words, the limit is zero, as we know:

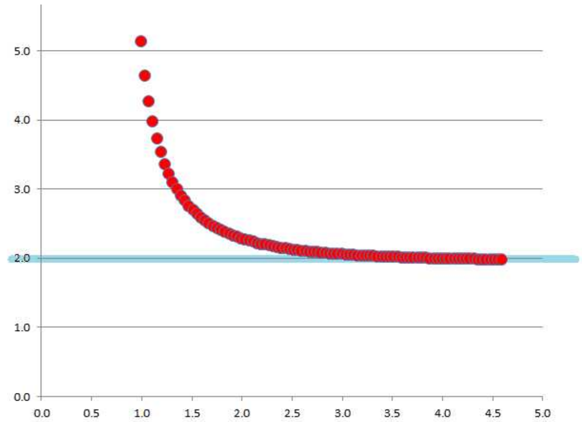
$$\frac{1}{n} \to 0 \quad \text{as } n \to \infty\,.$$

Example 1.5.4: computations

Evaluate the limit:

$$\lim_{n\to\infty} \frac{4n^2 - n + 2}{2n^2 - 1}\,.$$

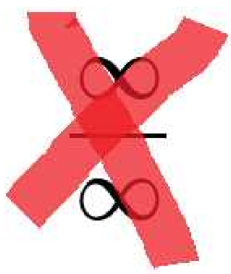
Plotting the graph suggests that the limit is $a = 2$:



Since the *last* operation is division, so we are supposed to use the *Quotient Rule* first. However, *both* the numerator and the denominator go to ∞. Then, the *Quotient Rule* is inapplicable!

An attempt to apply the *Quotient Rule* – against our better judgement – would result in an *indeterminate expression*:

$$\frac{4n^2 - n + 2}{2n^2 - 1} \quad \xrightarrow{???} \quad \frac{\infty}{\infty}\,.$$



Once again, we can't conclude that the limit doesn't exist; we've just failed to find the answer. The path out of this conundrum lies in algebra.

We need to get rid of the infinities in the fraction! How? The method that often works is as follows:

▶ Divide the numerator and denominator by an appropriate power of n.

Through trial and error, we determine that we should use n^2 :

$$\begin{aligned}
\frac{4n^2 - n + 2}{2n^2 - 1} &= \frac{(4n^2 - n + 2)/n^2}{(2n^2 - 1)/n^2} && \text{Divide by } n^2\,.\\
&= \frac{4 - \frac{1}{n} + \frac{2}{n^2}}{2 - \frac{1}{n^2}} && \text{Numerator and denominator consist of familiar sequences.}\\
&\to \frac{4 - 0 + 0}{2 - 0} && \text{Their limits are known.}\\
&= \frac{4}{2} && \text{SR allows us to evaluate these two limits.}\\
&= 2\,.
\end{aligned}$$

We only used the *Quotient Rule* at the very end, after ∞/∞ (the indeterminacy) was removed. In retrospect, we also see that the limit is the ratio of the leading coefficients. This is not a coincidence.

Exercise 1.5.5

Try to divide by x and x^3 instead.

The general method for finding such limits is given by the theorem below:

Theorem 1.5.6: Limits of Rational Functions at Infinity

Suppose we have a rational function f represented as a quotient of two polynomials of degrees p and q (the leading coefficients $a_p \neq 0$ and $b_q \neq 0$). Then the limit of the sequence defined by this function is the following:

$$\lim_{n\to\infty} \frac{a_pn^p + a_{p-1}n^{p-1} + ... + a_1n + a_0}{b_qn^q + b_{q-1}n^{q-1} + ... + b_1n + b_0} = \begin{cases} \pm\infty & \text{if } p > q \\ \frac{a_p}{b_p} & \text{if } p = q \\ 0 & \text{if } p < q \end{cases}$$

Proof.

The idea is, again, to *divide by the highest power*. If $p > q$, we have:

$$\frac{a_pn^p + a_{p-1}n^{p-1} + ... + a_1n + a_0}{b_qn^q + b_{q-1}n^{q-1} + ... + b_1n + b_0} = \frac{a_p + a_{p-1}n^{-1} + ... + a_1n^{-p+1} + a_0n^{-p}}{b_qn^{q-p} + b_{q-1}n^{q-p-1} + ... + b_1n^{1-p} + b_0n^{-p}} \to \frac{a_p + 0}{0} = \pm\infty\,.$$

If $p = q$, we have:

$$\frac{a_pn^p + a_{p-1}n^{p-1} + ... + a_1n + a_0}{b_qn^q + b_{q-1}n^{q-1} + ... + b_1n + b_0} = \frac{a_p + a_{p-1}n^{-1} + ... + a_1n^{-p+1} + a_0n^{-p}}{b_q + b_{q-1}n^{-1} + ... + b_1n^{1-p} + b_0n^{-p}} \to \frac{a_p + 0}{b_p + 0} = \frac{a_p}{b_p}\,.$$

If $p < q$, we have:

$$\frac{a_pn^p + a_{p-1}n^{p-1} + ... + a_1n + a_0}{b_qn^q + b_{q-1}n^{q-1} + ... + b_1n + b_0} = \frac{a_pn^{p-q} + a_{p-1}n^{p-q-1} + ... + a_1n^{1-q} + a_0n^{-q}}{b_q + b_{q-1}n^{-1} + ... + b_1n^{-q+1} + b_0n^{-q}} \to \frac{0}{b_q + 0} = 0\,.$$

This is the lesson we have re-learned:

- ▶ The long-term behavior of polynomials is determined by their leading terms.

We conclude that a larger degree polynomial will always "overpower" one with a lower degree: either the numerator takes the fraction to infinity or the denominator takes it to zero. Practically, we just drop the non-leading terms:

$$\lim_{n\to\infty} \frac{\boxed{a_pn^p} + a_{p-1}n^{p-1} + ... + a_1n + a_0}{\boxed{b_qn^q} + b_{q-1}n^{q-1} + ... + b_1n + b_0} = \lim_{n\to\infty} \frac{\boxed{a_pn^p}}{\boxed{b_qn^q}} = \frac{a_p}{b_q} \lim_{n\to\infty} n^{p-q}$$

This results in a power sequence, and we know its limit from the theorem in this chapter.

Example 1.5.7: rational sequences

With the theorem available, the algebraic trick we used in the last example isn't necessary anymore. In the following sequence, the degrees are equal; therefore, we have:

$$\frac{4n^3 - 2n + 2}{n^3 - 1} = \frac{4n}{1} = 4\,.$$

Example 1.5.8: non-rational sequences

The sequence below isn't rational:

$$\frac{1 + (-3)^n}{5^n}\,.$$

Therefore, the theorem doesn't apply. But the method does! First we notice that QR doesn't apply:

$$\begin{array}{rll}
y_n & = \dfrac{1+(-3)^n}{5^n} & \text{The numerator diverges – DEAD END!}\\
& = \dfrac{1}{5^n}+\dfrac{(-3)^n}{5^n} & \text{The way out is always algebra: We will try SR.}\\
& = \left(\dfrac{1}{5}\right)^n+\left(\dfrac{-3}{5}\right)^n & \text{We use properties of exponents and discover geometric progressions.}\\
& \quad\ \downarrow \qquad\quad\ \downarrow & \text{Take the limits.}\\
& \quad\ 0 \qquad\quad\ 0 & \text{Because the ratios are within } (-1,1)\,.\\
& \to \quad 0 & \text{According to SR.}
\end{array}$$

These are two geometric progressions with the ratios: $r = 1/5,\ -3/5$, that satisfy $|r| < 0$. Meanwhile, our application of the *Sum Rule* was justified by the fact that the two limits exist.

Example 1.5.9: non-rational sequences

The sequence below isn't rational:

$$\frac{(-3)^n+1}{5^n+1}\,.$$

Therefore, the theorem doesn't apply. But the trick we used to deal with ∞/∞ does. We divide both numerator and denominator by the "strongest" sequence present, 5^n, then simplify:

$$\frac{(-3)^n+1}{5^n+1}=\frac{(-3/5)^n+1/5^n}{1+1/5^n}\to\frac{0+0}{1+0}=0\,.$$

Exercise 1.5.10

Find the limit of the composition of $f(x) = \text{sign}(x)$ and the sequence x_n given by (a) $1/n$, (b) $-1/n$, (c) $(-1)^n/n$.

What is infinity?

The plus (or minus) infinity is identified with the collection of all sequences approaching this infinity. In other words, the following identity is read in both directions:

$$\lim_{n\to+\infty} a_n = +\infty\,.$$

Now, does it make sense to do any algebra with the infinities? This seems to make sense:

$$\infty+\infty=\infty\,.$$

But this doesn't:

$$\infty-\infty=???$$

We will have to see when the algebra with those infinite limits makes sense.

The *Algebraic Rules of Limits* above have exceptions; we can imagine that one or both of the sequences approach infinity or that the limit in the denominator is 0.

First, if two sequences approach the *same* infinity, then so does their sum. Similar for the product:

Theorem 1.5.11: Algebra of Infinite Limits of Sequences

Suppose

- $a_n \to \pm\infty$, *and*
- $b_n \to \pm\infty$.

Then we have:

$$\begin{array}{lll} \text{SR:} & a_n + b_n & \to \pm\infty \\ \text{PR:} & a_n \cdot b_n & \to +\infty \end{array}$$

It is just as important what is NOT here:

$$\begin{array}{lll} \text{DR?} & a_n - b_n & \to ??? \\ \text{QR?} & a_n / b_n & \to ??? \end{array}$$

Example 1.5.12: infinite limits

The Sum Rule and the Product Rule, as stated in the last section, are inapplicable because neither of the two limits exists, but the last theorem applies:

$$n^2 + 2^n \to \infty \quad \text{and} \quad n^2 \cdot 2^n \to \infty\,.$$

Second, if one limit is infinite and the other is not, the sum or product is infinite. Furthermore, we have the following:

Theorem 1.5.13: Algebra of Finite and Infinite Limits of Sequences

Suppose

- $a_n \to a$, *and*
- $b_n \to \pm\infty$.

Then we have:

$$\begin{array}{ll|lll} \text{SR:} & a_n + b_n \to \pm\infty & \text{CMR:} & c \cdot b_n \to \pm\infty & \text{for any real } c > 0 \\ \text{PR:} & a_n \cdot b_n \to \pm\infty & \text{QR:} & a_n/b_n \to 0, & b_n/a_n \to \pm\infty \\ & \text{provided } a > 0 & & & \text{provided } a > 0 \end{array}$$

Example 1.5.14: finite and infinite limits

The Sum Rule and the Product Rule, as stated in the last section, are inapplicable because one of the two limits doesn't exist, but the last theorem applies:

$$\frac{1}{n} + 2^n \to \infty \quad \text{and} \quad \frac{n}{n+1} \cdot 2^n \to \infty\,.$$

Exercise 1.5.15

Consider the cases of the theorem when $a < 0$ and $a = 0$.

Justified by these theorems, we make this informal list for the *algebra of infinities*:

$$\begin{array}{llll} \text{number} & + & (+\infty) & = +\infty \\ \text{number} & + & (-\infty) & = -\infty \\ +\infty & + & (+\infty) & = +\infty \\ -\infty & + & (-\infty) & = -\infty \\ \text{number} & / & (\pm\infty) & = 0 \end{array}$$

Once again, this isn't on the list:

$$\begin{array}{cccl} +\infty & - & (+\infty) & = ??? \\ +\infty & / & (+\infty) & = ??? \end{array}$$

Example 1.5.16: $\infty - \infty$

There is no $\infty - \infty$ (just as there is no ∞/∞). Why not?

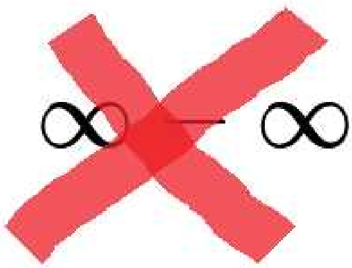
Behind each ∞, there must be a sequence approaching ∞! However, the outcome is ambiguous. On the one hand we have:

$$a_n = n \to +\infty,\ b_n = -n \to +\infty \implies a_n - b_n = 0 \to 0\,;$$

on the other:

$$a_n = n^2 \to +\infty,\ b_n = -n \to +\infty \implies a_n - b_n = n^2 - n \to +\infty\,,$$

by *Limits of Polynomials*. Two seemingly legitimate answers for the same expression, $\infty - \infty$...



We have another indeterminate expression!

Example 1.5.17: 1^∞

What about this limit?

$$1^\infty\,.$$

Of course, if we keep multiplying 1 by itself, we will always have 1. However, what if 1 is also the limit of a sequence? Do we have:

$$1^\infty = \lim_{n\to\infty} a_n^{b_n}\,,$$

for some (or any?) sequences:

$$\lim_{n\to\infty} a_n = 1, \quad \lim_{n\to\infty} b_n = \infty\,?$$

Let's try

$$a_n = 1 + \frac{1}{n} \ \text{ and } \ b_n = n\,.$$

We will show at the end of the chapter that the limit exists:

$$\left(1 + \frac{1}{n}\right)^n\,.$$

So, 1^∞ can be another indeterminate expression!

1.6. More properties of limits of sequences

Every sequence *is* a function; it just happens to have a special kind of domain. Then, why do they deserve a special attention?

A sequence doesn't produce a function easily when it is defined *recursively*.

Example 1.6.1: recursive limits

Recursive definitions are very common; even the simplest banking requires one to use them:

1. If you say that you will contribute \$2000 every year, you are stating that the next year's balance will be \$2000 higher than the last:

$$a_{n+1} = a_n + 2000\,.$$

2. If you say that your bank will pay 5% per year, you are stating that the next year's balance will be 1.05 times higher than the last:

$$b_{n+1} = b_n \cdot 1.05\,.$$

Of course, we then *derive* the nth-term formulas for these sequences:

1. repeated deposits starting from 0:

$$a_n = 2000 \cdot n\,;$$

2. compounded interest starting from \$2000:

$$b_n = 2000 \cdot 1.05^n\,.$$

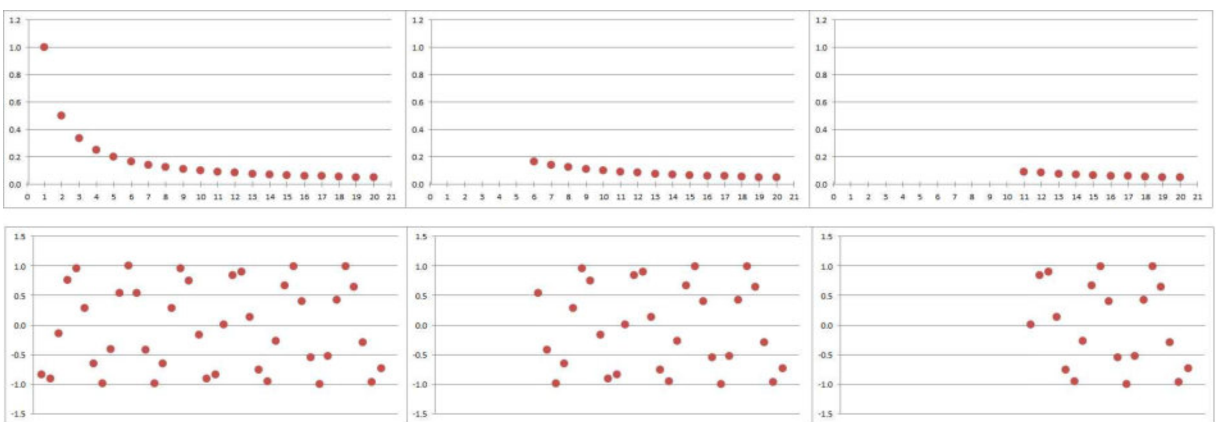
Then it is clear from these formulas that the limits are infinite. However, what if we carry out both of the strategies? The recursive formula is still very clear:

$$c_{n+1} = (c_n + 2000) \cdot 1.05\,.$$

But there is no nth-term formula! How do we even prove that the limit is infinite? Indirectly, by *comparing* c_n with either a_n or b_n.

This section is about *indirect* proofs of convergence or divergence.

First, compare these sequences:



The observation is very simple but useful: Only the *tail* of the sequence matters for convergence!

Theorem 1.6.2: Truncation Principle

A sequence is convergent if and only if all of its truncations are convergent.

In other words, we have:

$$\boxed{a_n\Big|_{n=p,p+1,...} \to a \iff a_n\Big|_{n=q,q+1,...} \to a}$$

for any integer p and q.

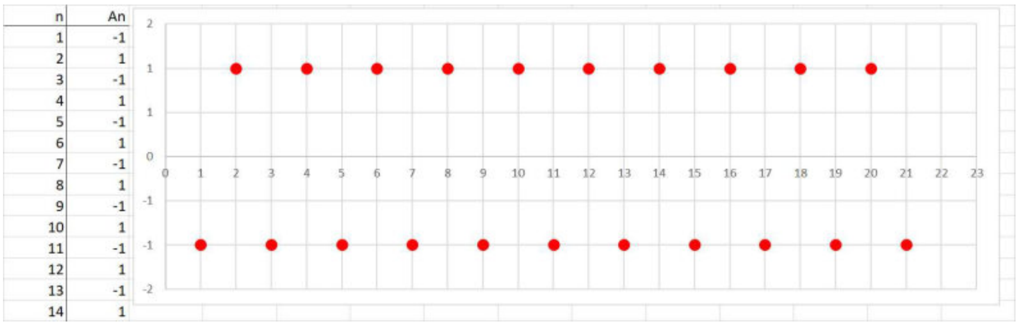
Truncation amounts to a *restriction of the domain* of a function (seen in Volume 1, Chapter 1PC-4). Using the other notation for limits, we have:

$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} a_n\Big|_{n\geq p} = \lim_{n\to\infty} a_n\Big|_{n\geq q}$$

Exercise 1.6.3

Prove the theorem.

One can see other important ways to restrict the domain of a sequence. For example, the alternating sequence diverges, but when restricted to the even or the odd numbers, it produces two *convergent* sequences:



n	An
1	-1
2	1
3	-1
4	1
5	-1
6	1
7	-1
8	1
9	-1
10	1
11	-1
12	1
13	-1
14	1

This convenient idea is summarized as follows:

Definition 1.6.4: subsequence

A sequence restricted to an infinite subset of the integers is called its *subsequence.*

For example, these are some subsequences of the alternating sequence $(-1)^n$:

$$(-1)^{2n},\ (-1)^{3n},\ (-1)^{n^2}.$$

They behave as expected:

Theorem 1.6.5: Limit of Subsequence

A subsequence of any convergent sequence is also convergent, and the limit is the same.

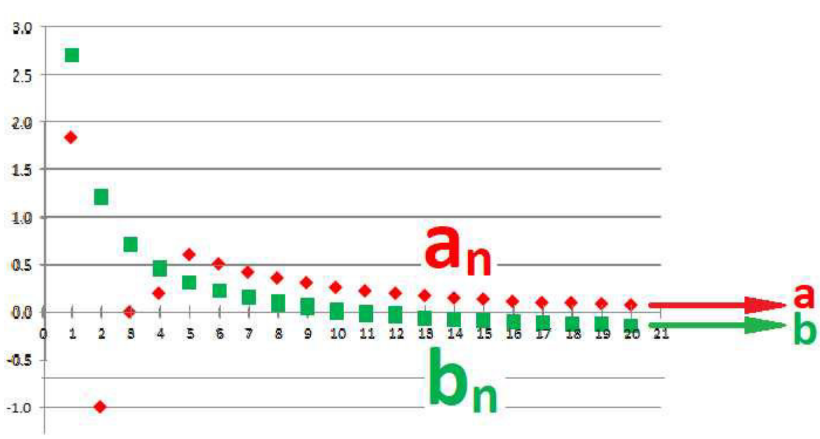
The converse is untrue: $(-1)^n$ is divergent, but $(-1)^{2n} = 1$ is convergent.

Non-strict inequalities between sequences, such as:

$$a \leftarrow a_n \geq b_n \to b,$$

are preserved under limits:

$$a \geq b\,.$$



Theorem 1.6.6: Comparison Test for Limits of Sequences

If $a_n \geq b_n$ for all n greater than some N, then

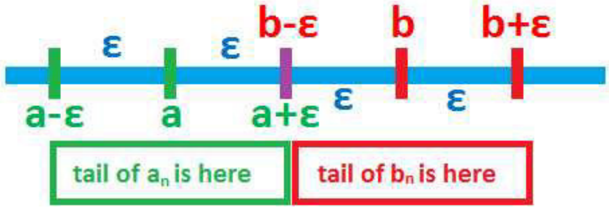
$$\lim_{n\to\infty} a_n \geq \lim_{n\to\infty} b_n\,,$$

provided the sequences converge.

Proof.

The proof is by contradiction. Suppose a and b are the limits of a_n and b_n respectively and suppose also $a < b$.

The geometry of the proof is clear: We want to separate the two horizontal lines representing the two limits by two non-overlapping bands, just as last time. Then, if narrow enough, the tails of the "larger" sequence would have to fit the "smaller" band. These bands correspond to two intervals around those two limits. In order for them to not intersect, their width (that's 2ε!) should be less than half the distance between the two numbers.



Let

$$\varepsilon = \frac{b-a}{2}\,.$$

Then, what we are going to use at the end is

$$a + \varepsilon = b - \varepsilon\,.$$

Now, we use the definition for a and b as limits:

1. There exists a natural number N such that for every natural number $n > L$, we have
$$|a_n - a| < \varepsilon\,.$$
2. There exists a natural number K such that for every natural number $n > M$, we have
$$|b_n - b| < \varepsilon\,.$$

In order to combine the two statements, we need them to be satisfied for the same values of n. Let

$$N = \max\{L, M\}\,.$$

Then,

1. For every number $n > N$, we have
$$|a_n - a| < \varepsilon\ , \ \text{ or } \ a - \varepsilon < a_n < a + \varepsilon\,.$$
2. For every number $n > N$, we have
$$|b_n - b| < \varepsilon, \ \text{ or } \ b - \varepsilon < b_n < b + \varepsilon\,.$$

Taking one from either of the two pairs of inequalities, we have:

$$a_n < a + \varepsilon = b - \varepsilon < b_n\,.$$

A contradiction.

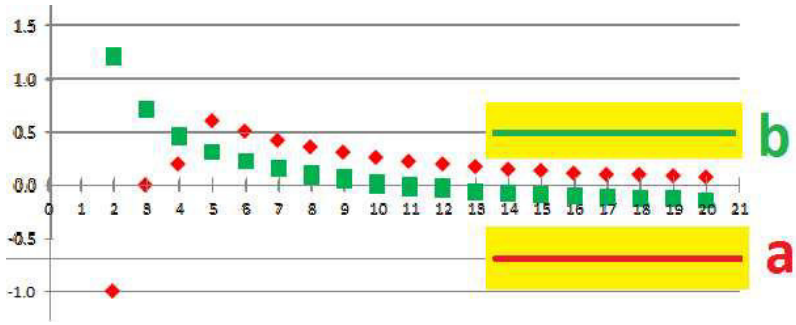
Warning!

The theorem does not claim convergence.

Exercise 1.6.7

Show that replacing the non-strict inequality, $a_n \geq b_n$, with a strict one, $a_n > b_n$, won't produce a strict inequality in the conclusion of the theorem.

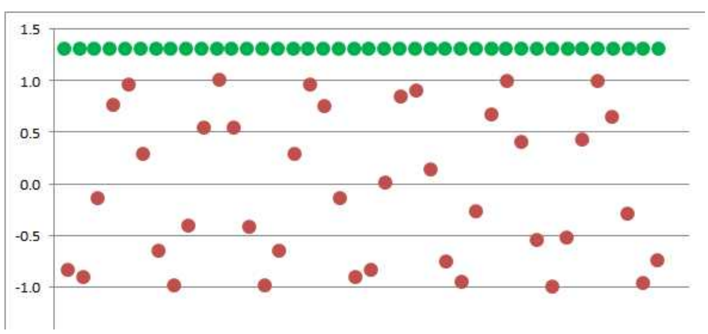
The situation is similar to that of the *Uniqueness Theorem*: If the opposite inequality were to hold, we could find two bands to contain the two sequences' tails. Then the original inequality would fail:



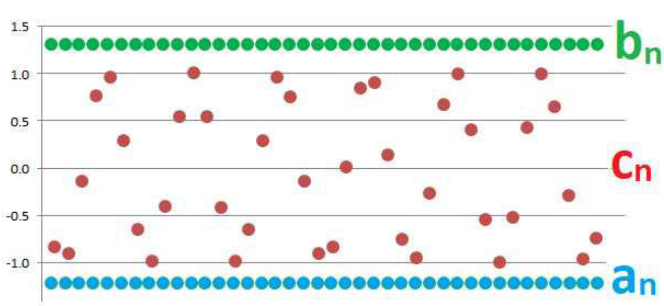
This is the summary of the theorem:

$$\begin{array}{cccc} & a_n & \geq & b_n \\ & \downarrow & & \downarrow \\ & a & & b \\ \Longrightarrow & & \geq & \end{array}$$

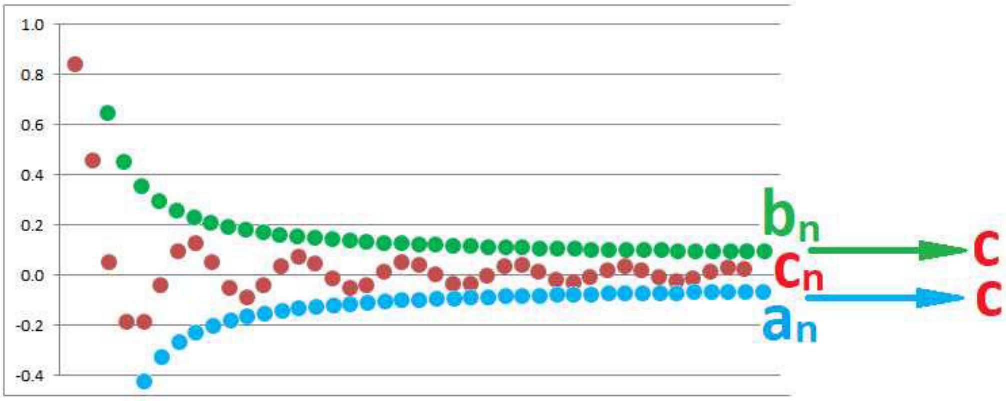
From the inequality in the theorem, we can't conclude anything about the existence of the limit:



Let's try having an inequality on *either side* of the sequence:



However, they'd have to work together in order to control the sequence they bound! It is called a *squeeze*:



If we can squeeze the sequence under investigation between two familiar sequences, we might be able to say something about its limit. Some further requirements will be necessary:

Theorem 1.6.8: Squeeze Theorem for Sequences

If the values of a sequence lie between those of two sequences with the same limit, then its limit also exists and is equal to that number.

In other words, we have:

$$a_n \leq c_n \leq b_n \quad \text{for all } n > N\,,$$

for some N, and

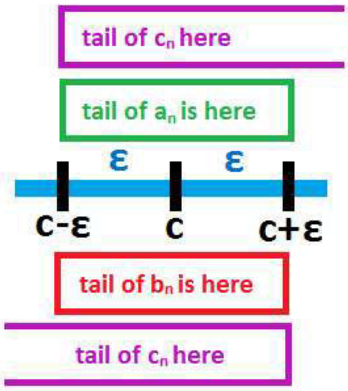
$$\lim_{n\to\infty} a_n = \lim_{n\to\infty} b_n = c\,,$$

then the sequence c_n converges, and

$$\lim_{n\to\infty} c_n = c\,.$$

Proof.

The geometry of the proof is shown below:



Suppose $\varepsilon > 0$ is given. As we know, we have for all n larger than some N:

$$c - \varepsilon < a_n < c + \varepsilon \quad \text{and} \quad c - \varepsilon < b_n < c + \varepsilon\,.$$

Then we have:

$$c - \varepsilon < a_n \leq c_n \leq b_n < c + \varepsilon\,.$$

Warning!

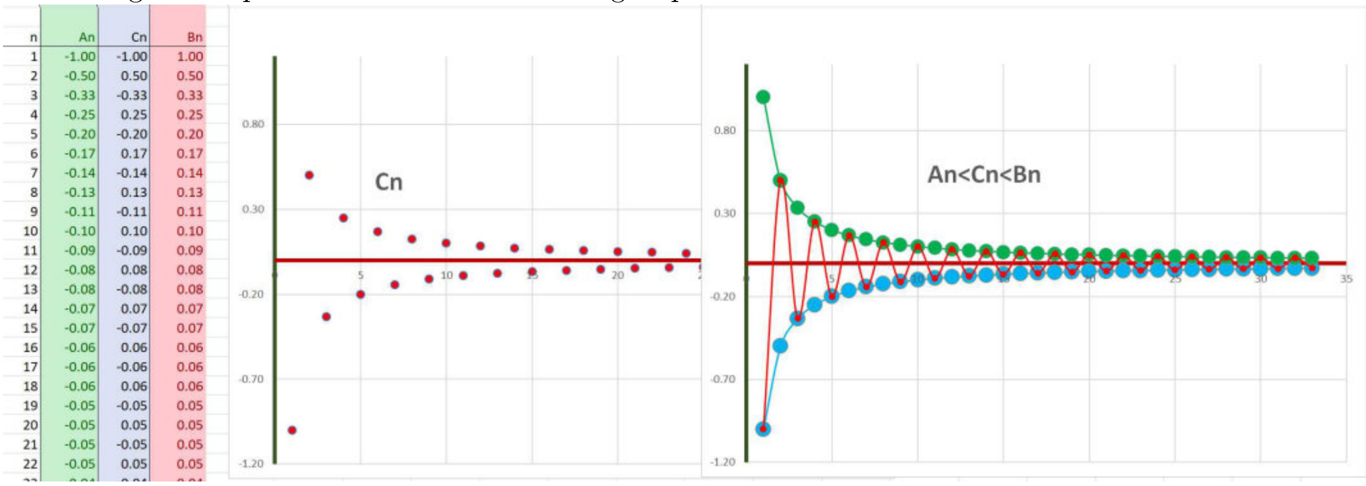
The theorem does claim convergence this time.

Example 1.6.9: alternating reciprocals

Sometimes the choice of the squeeze is obvious. Let's imagine that we see this sequence for the first time:

$$c_n = \frac{(-1)^n}{n}\,.$$

Examining the sequence reveals two bounding sequences:



n	An	Cn	Bn
1	-1.00	-1.00	1.00
2	-0.50	0.50	0.50
3	-0.33	-0.33	0.33
4	-0.25	0.25	0.25
5	-0.20	-0.20	0.20
6	-0.17	0.17	0.17
7	-0.14	-0.14	0.14
8	-0.13	0.13	0.13
9	-0.11	-0.11	0.11
10	-0.10	0.10	0.10
11	-0.09	-0.09	0.09
12	-0.08	0.08	0.08
13	-0.08	-0.08	0.08
14	-0.07	0.07	0.07
15	-0.07	-0.07	0.07
16	-0.06	0.06	0.06
17	-0.06	-0.06	0.06
18	-0.06	0.06	0.06
19	-0.05	-0.05	0.05
20	-0.05	0.05	0.05
21	-0.05	-0.05	0.05
22	-0.05	0.05	0.05

We have:

$$-\frac{1}{n} \le \frac{(-1)^n}{n} \le \frac{1}{n}\,.$$

Now, since both $a_n = -\dfrac{1}{n}$ and $b_n = \dfrac{1}{n}$ go to 0, by the *Squeeze Theorem*, so does $c_n = \dfrac{(-1)^n}{n}$.

The example shows that the absolute value is a good tool for constructing a squeeze. We apply this idea to all sequences:

Corollary 1.6.10: Limit of Absolute Value

A sequence converges to zero if and only if its absolute value does.

In other words, we have:

$$a_n \to 0 \iff |a_n| \to 0\,.$$

Exercise 1.6.11

Prove the corollary.

Exercise 1.6.12

What if it's not 0?

Example 1.6.13: diminishing oscillations

Let's find the limit,

$$\lim_{n\to\infty} \frac{1}{n}\sin n\,.$$

It cannot be computed by the *Product Rule* because $\lim\limits_{n\to\infty} \sin n$ does not exist. Let's try a squeeze. This is what we know from trigonometry:

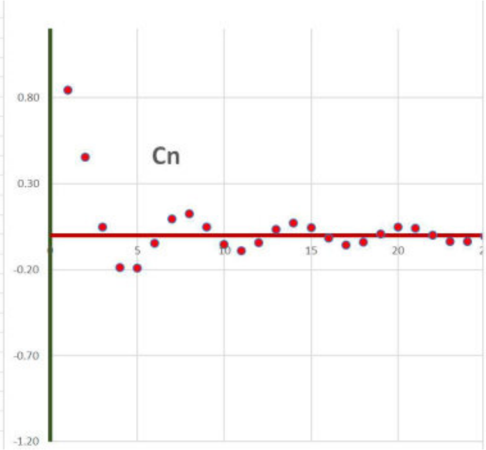
$$-1 \le \sin n \le 1\,.$$

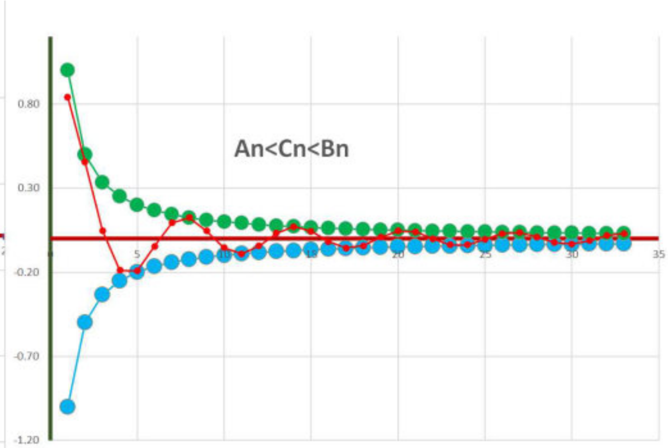
However, this squeeze proves nothing about the limit of our sequence!

Let's try another squeeze:

$$-\left|\frac{1}{n}\right| \le \frac{1}{n}\sin n \le \left|\frac{1}{n}\right| .$$

n	An	Cn	Bn
1	-1.00	0.84	1.00
2	-0.50	0.45	0.50
3	-0.33	0.05	0.33
4	-0.25	-0.19	0.25
5	-0.20	-0.19	0.20
6	-0.17	-0.05	0.17
7	-0.14	0.09	0.14
8	-0.13	0.12	0.13
9	-0.11	0.05	0.11
10	-0.10	-0.05	0.10
11	-0.09	-0.09	0.09
12	-0.08	-0.04	0.08
13	-0.08	0.03	0.08
14	-0.07	0.07	0.07
15	-0.07	0.04	0.07
16	-0.06	-0.02	0.06
17	-0.06	-0.06	0.06
18	-0.06	-0.04	0.06
19	-0.05	0.01	0.05
20	-0.05	0.05	0.05
21	-0.05	0.04	0.05
22	-0.05	0.00	0.05





Now, since

$$\lim_{n\to\infty}\left(-\frac{1}{n}\right) = \lim_{n\to\infty}\frac{1}{n} = 0, \, ,$$

from the *Squeeze Theorem*, we conclude:

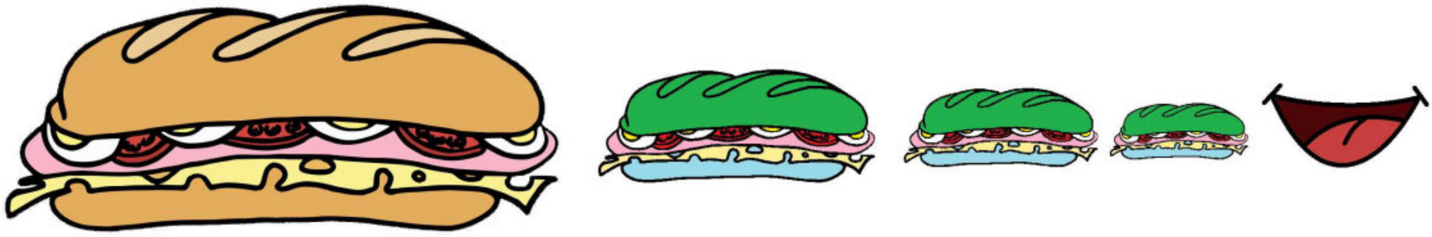
$$\lim_{n\to\infty}\frac{1}{n}\sin n = 0 \, .$$

Exercise 1.6.14

Suppose a_n and b_n are convergent. Prove that $\max\{a_n, b_n\}$ and $\min\{a_n, b_n\}$ are also convergent. Hint: Start with the case $\lim a_n > \lim b_n$.

The squeeze theorem is also known as the *Two Policemen Theorem*: If two policemen are escorting a prisoner handcuffed between them, and both officers go to the same(!) police station, then – in spite of some freedom the handcuffs allow – the prisoner will also end up in that station:

Another name is the *Sandwich Theorem*. It is, once again, about control. A sandwich can be a messy affair: ham, cheese, lettuce, etc. One won't want to touch that and instead takes control of the contents by keeping them between the two buns. He then brings the two to his mouth and the rest of the sandwich along with them!



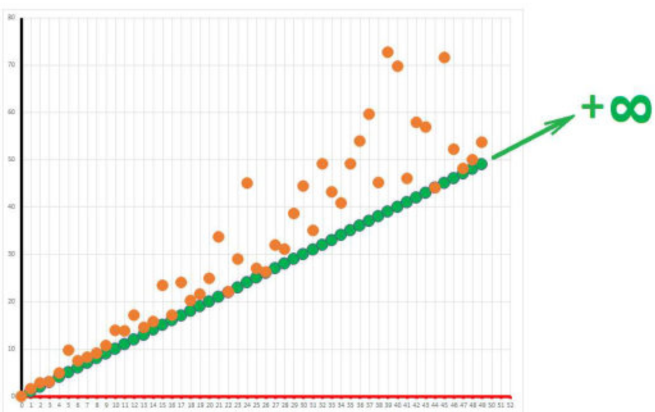
We summarize the theorem by pointing out the five hypotheses and the two conclusions:

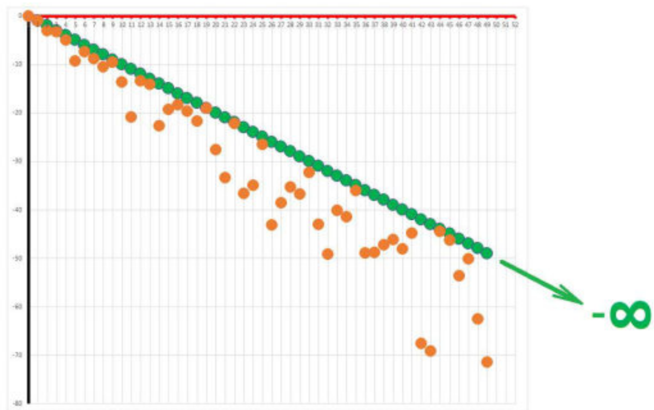
$$\begin{array}{cccccc}
 & a_n & \leq_{(1)} & c_n & \leq_{(2)} & b_n \\
 & \downarrow_{(3)} & & & & \downarrow_{(4)} \\
 & a & = & =_{(5)} & = & b \\
\Longrightarrow & & & \downarrow_{(6)} & & \\
 & & & a =_{(7)} b & &
\end{array}$$

Exercise 1.6.15

List these statements.

To make conclusions about *divergence to infinity*, we only need to control it from *one*, but the right, side:





The smaller sequence will push the larger to infinity, and the larger sequence will push the smaller to negative infinity. Below is an analog of the Squeeze Theorem for infinite limits:

Theorem 1.6.16: Push Out for Limits of Sequences

1. *If the values of a sequence lie above those of a sequence that diverges to positive infinity, then so does this sequence.*
2. *If the values of a sequence lie below those of a sequence that diverges to negative infinity, then so does this sequence.*

In other words, if $a_n \geq b_n$ for all n greater than some N, then we have:

$$\begin{array}{ll}
1. & \lim_{n\to\infty} a_n = -\infty \quad \Longrightarrow \quad \lim_{n\to\infty} b_n = -\infty\,. \\
2. & \lim_{n\to\infty} a_n = +\infty \quad \Longleftarrow \quad \lim_{n\to\infty} b_n = +\infty\,.
\end{array}$$

Exercise 1.6.17

Prove the theorem.

Exercise 1.6.18

Suppose a sequence is defined recursively by

$$a_{n+1} = 2a_n + 1 \text{ with } a_0 = 1\,.$$

Does the sequence converge or diverge?

This is the summary of the theorem:

$$
\begin{array}{ccc|ccc}
 & a_n \quad \leq & c_n & & c_n \quad \leq & b_n \\
 & \downarrow & & & & \downarrow \\
 & +\infty & & & & -\infty \\
\Longrightarrow & & \downarrow & \Longrightarrow & \downarrow & \\
 & & +\infty & & -\infty &
\end{array}
$$

1.7. Theorems of Analysis

The theorems in this section will be used to prove new theorems. It can be skipped on the first reading.

We accept the following fundamental result without proof:

Theorem 1.7.1: Monotone Convergence Theorem

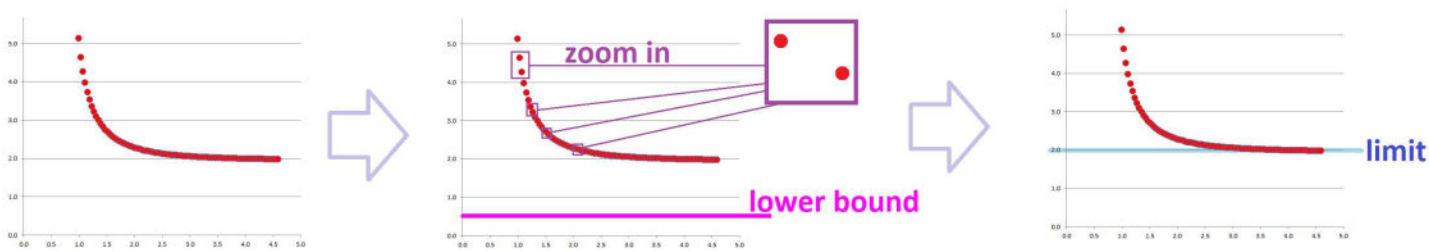
Every bounded and monotonic sequence is convergent.

In other words, if a sequence a_n is

- *either increasing, $a_n \leq a_{n+1}$ for all n, or decreasing, $a_n \geq a_{n+1}$ for all n, and*
- *bounded, $|a_n| \leq Q$ for some number Q,*

then it has a limit.

Simply knowing these two facts proves that there is a limit:



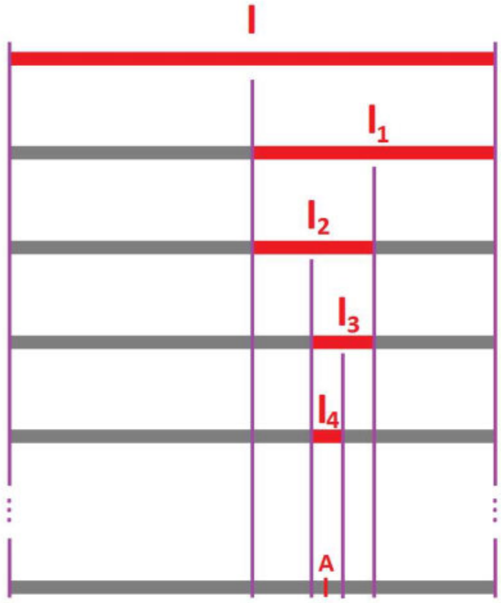
However, there is no information about the value of the limit.

The result is also known as the *Completeness Property of Real Numbers*.

Another result with regard to the geometry of real numbers is about a shrinking sequence of intervals:

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset \ldots$$

It is illustrated below:



The conclusion suggested by the picture is presented below:

Theorem 1.7.2: Nested Intervals Theorem

1. *A sequence of nested closed intervals has a non-empty intersection, i.e., if we have two sequences of numbers a_n and b_n that satisfy*

$$a_1 \leq a_2 \leq ... \leq a_n \leq ... \leq b_n \leq ... \leq b_2 \leq b_1 \,,$$

then they both converge,

$$a_n \to a, \;\; b_n \to b \,,$$

and

$$\bigcap_{n=1}^{\infty} [a_n, b_n] = [a, b] \,.$$

2. *If, moreover,*

$$b_n - a_n \to 0 \,,$$

then

$$\bigcap_{n=1}^{\infty} [a_n, b_n] = \{a\} = \{b\} \,.$$

Proof.

For part (1), observe that a point x belongs to the intersection if and only if it satisfies:

$$a_n \leq x \leq b_m, \quad \text{for all } n, m \,.$$

Meanwhile, the sequences converge by the *Monotone Convergence Theorem*. Therefore,

$$a \leq x \leq b$$

by the *Comparison Theorem*.

For part (2), consider:

$$0 = \lim_{n\to\infty} (b_n - a_n) = \lim_{n\to\infty} b_n - \lim_{n\to\infty} a_n = b - a \,,$$

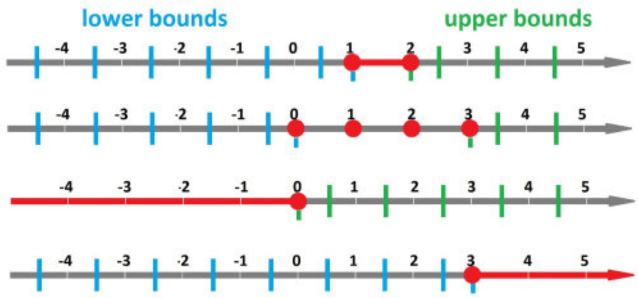
by the *Sum Rule*. We then conclude that $a = b$.

We have indeed a "nested" sequence of intervals

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset \dots,$$

with a single point, say A, in common.

Let's recall some definitions (seen in Volume 1, Chapter 1PC-2).



Definition 1.7.3: upper and lower bounds of set

Suppose S is a set of real numbers.

- An *upper bound* of S is any number M that satisfies:

$$x \le M \text{ for any } x \text{ in } S.$$

- A *lower bound* of S is any number m that satisfies:

$$x \ge m \text{ for any } x \text{ in } S.$$

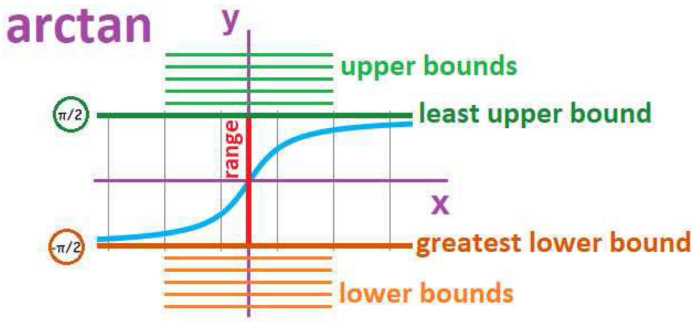
For $S = [0, 1]$, any number $M \ge 1$ is its upper bound. However, these sets have no upper bounds:

$$(-\infty, +\infty),\ [0, +\infty),\ \{0, 1, 2, 3, \dots\}.$$

We now take this to the next level:

Definition 1.7.4: sets bounded below and above

A set that has an upper bound is called *bounded above*, and a set that has a lower bound is called *bounded below*. A set that has both upper and lower bounds is called *bounded*; otherwise, it's *unbounded*.



Definition 1.7.5: supremum and infimum

Suppose S is a set.

- An upper bound for which there is no smaller upper bound is called the *least upper bound*; it is also called the *supremum*. It is denoted as follows:

$$\boxed{\sup S}$$

- A lower bound for which there is no larger lower bound is called the *greatest lower bound*; it is also called the *infimum*. It is denoted as follows:

$$\boxed{\inf S}$$

Below we justify the usage of "the":

Theorem 1.7.6: Uniqueness of Supremum and Infimum

- *For a given set, there can be only one least upper bound.*
- *For a given set, there can be only one greatest lower bound.*

Proof.

Thus, $M = \sup S$ means that:

1. M is an upper bound of S.
2. If M' is another upper bound of S, then $M' \geq M$.

Now, if we have another $M' = \sup S$, then:

1. M' is an upper bound of S.
2. If M is another upper bound of S, then $M \geq M'$.

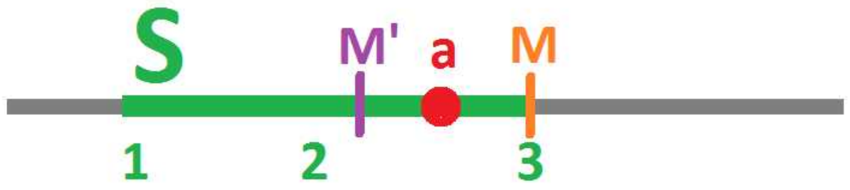
Therefore, $M = M'$.

Example 1.7.7: least upper bound

For the following sets, the least upper bound is $M = 3$:

- $S = \{1, 2, 3\}$
- $S = [1, 3]$
- $S = (1, 3)$

The proof for the last one is as follows. Suppose M' is an upper bound with $1 < M' < 3$. Let's choose $a = \dfrac{3 + M'}{2}$. But a belongs to S! Therefore, M' isn't an upper bound:



What if we limit S to the *rational* numbers only in $(1, 3)$? Then $a = \dfrac{3 + M'}{2}$ won't belong to S when M' is irrational. The proof fails.

We know now that there is at most one such number. What about at least?

Theorem 1.7.8: Existence of Supremum and Infimum

- *Any bounded above set has a least upper bound.*
- *Any bounded below set has a greatest lower bound.*

Proof.

The idea of the proof is to construct nested intervals with the right-end points being upper bounds. What should be the left-end points?

Suppose a set S is given. Then:

1. Let U be the set of all upper bounds of S.
2. Let L be the set of all lower bounds of U.

Since S is bounded, we have:

$$U \neq \varnothing\,.$$

Now, if S is a single point, we are done. If not, we have x, y in S such that $x < y$. Therefore, x belongs to L, and

$$L \neq \varnothing\,.$$

We start with:

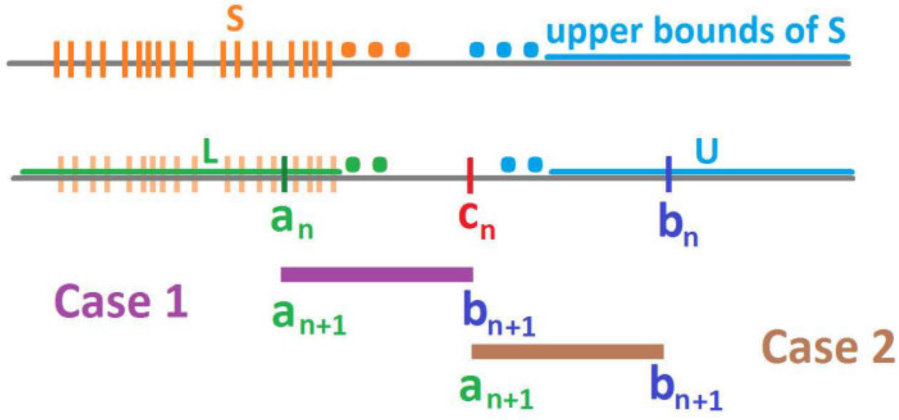
- ▶ a_1 is any element of L, and b_1 is any element of U.

Suppose inductively that we have constructed two sequences of numbers

$$a_i,\ b_i, \quad i = 1, 2, 3..., n\,,$$

such that:

1. a_i is in L and b_i is in U.
2. $a_n \leq ... \leq a_1 \leq b_1 \leq ... \leq b_n$.
3. $b_i - a_i \leq \frac{1}{2^{i-1}}(b_1 - a_1)$.



We continue with the inductive step: Let

$$c = \frac{1}{2}(b_n - a_n)\,.$$

We have two cases.

Case 1: c belongs to U. Then choose

$$a_{n+1} = a_n \quad \text{and} \quad b_{n+1} = c\,.$$

Then,

$$a_{n+1} = a_n \text{ in } L \quad \text{and} \quad b_{n+1} = c \text{ in } U\,.$$

Case 2: c belongs to L. Then choose

$$a_{n+1} = c \quad \text{and} \quad b_{n+1} = b_n\,.$$

Then,

$$a_{n+1} = c \text{ in } L \quad \text{and} \quad b_{n+1} = b_n \text{ in } U\,.$$

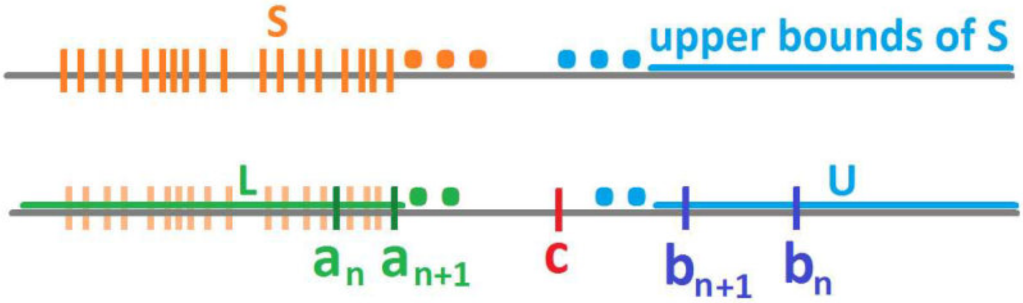
Furthermore,

$$b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n) \le \frac{1}{2}\frac{1}{2^{n-1}}(b_1 - a_1) = \frac{1}{2^n}(b_1 - a_1)\,.$$

Thus, all the conditions are satisfied, and our sequence of nested intervals has been inductively built. We apply the *Nested Intervals Theorem* and conclude that

$$a_n \to d \leftarrow b_n\,.$$

Why is c a least upper bound of S?



First, suppose c is *not* an upper bound. Then there is x in S with $x > c$. If we choose $\varepsilon = x - c$, then from $b_n \to c$ we conclude that $b_n < x$ for all $n > N$ for some N. This contradicts the assumption that b_n in U.

Second, suppose c is *not* a least upper bound. Then there is an upper bound $y < c$. If we choose $\varepsilon = c - y$, then from $a_n \to c$ we conclude that $a_n > y$ for all $n > N$ for some N. This contradicts the assumption that a_n in L.

Below we describe what it means for a set to be an *interval*:

Theorem 1.7.9: Intermediate Point Theorem

A subset J of the reals is an interval or a single point if and only if it contains all of its intermediate points.

In other words, we have:

$$y_1,\ y_2 \ \text{ in } \ J \quad \texttt{AND} \quad y_1 < c < y_2 \implies c \ \text{ in } \ J\,.$$

Proof.

The "if" part is obvious. Now assume that the condition is satisfied for set J. Suppose also that J is bounded. Then these exist by the *Existence of Supremum Theorem*:

$$a = \inf S,\ b = \sup J\,.$$

Note that these might not belong to J. However, if c satisfies $a \le c \le b$, then there are these two:

1. y_1 in J such that $a < y_1 < c$, and
2. y_2 in J such that $c < y_2 < b$.

By the property, then we have: c in J. Therefore, J is an interval with a, b as its end-points.

Exercise 1.7.10

Prove the theorem for the unbounded case.

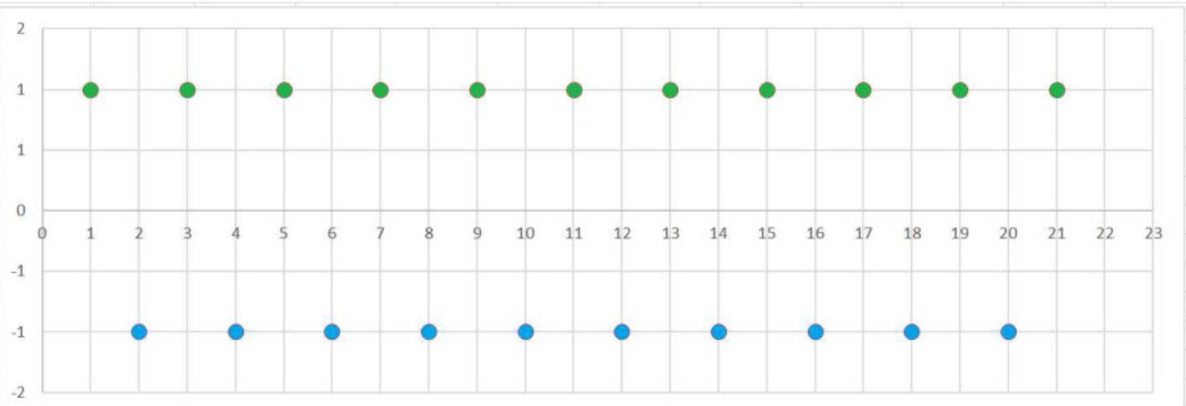
Example 1.7.11: subsequences

The converse of the *Monotone Convergence Theorem* above is, of course, false: The alternating sequence $c_n = (-1)^n$ is both bounded (it lies within $[-1, 1]$) and divergent. We, however, can't help noticing that it is made of two *convergent* sequences:

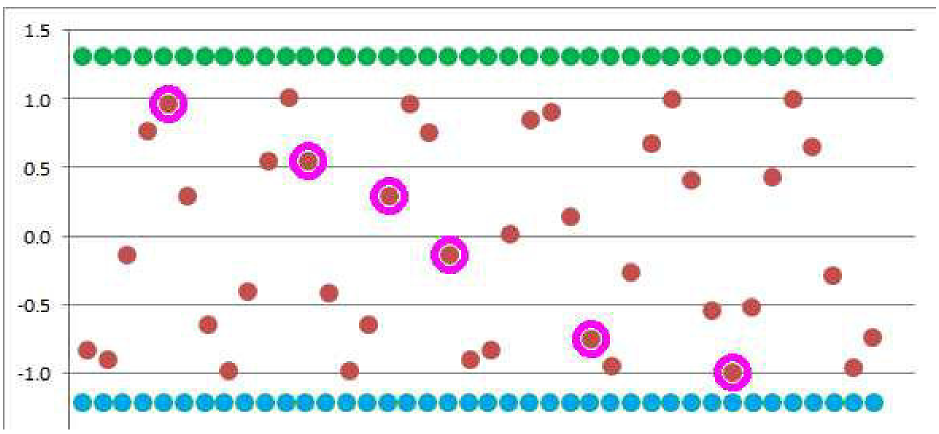
n even	$c_n = 1$	a_n
n odd	$c_n = -1$	b_n

These two are *subsequences* of c_n.

n	An
1	1
2	-1
3	1
4	-1
5	1
6	-1
7	1
8	-1
9	1
10	-1
11	1
12	-1
13	1



We can always extract a subsequence that is either increasing or decreasing:



The reason is that there are always infinitely many terms either below or above every term of the sequence. We have reached the following conclusion:

Theorem 1.7.12: Bolzano-Weierstrass Theorem

Every bounded sequence has a convergent subsequence.

Proof.

Suppose x_n is such a sequence. Then, it is contained in some interval $[a, b]$. The first part of the construction is to cut consecutive intervals in half and pick the half that contains infinitely many elements of the set $\{x_n : \ n = 1, 2, 3...\}$.

Similar to the previous proofs, we assume that we have already constructed sequences:

$$a_i, \ b_i, \ i = 1, 2, 3..., n \, ,$$

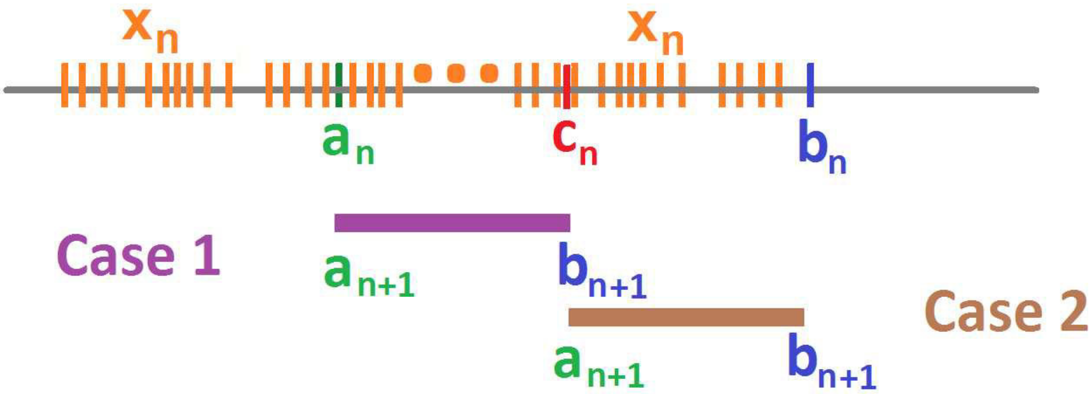
such that:

1. The interval $[a_i, b_i]$ contains infinitely many elements of $\{x_n : \ n = 1, 2, 3...\}$.
2. $a_n \le ... \le a_1 \le b_1 \le ... \le b_n$.
3. $b_i - a_i \le \dfrac{1}{2^{i-1}}(b_1 - a_1)$.

We continue with the inductive step: Let

$$c = \frac{1}{2}(b_n - a_n)\,.$$

We have two cases.



Case 1: Interval $[a_n, c]$ contains infinitely many elements of $\{x_n : \ n = 1, 2, 3...\}$. Then choose

$$a_{n+1} = a_n \quad \text{and} \quad b_{n+1} = c\,.$$

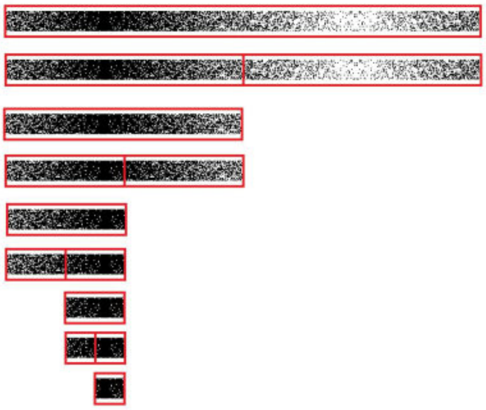
Case 2: Interval $[a_n, c]$ does not contain infinitely many elements of $\{x_n : \ n = 1, 2, 3...\}$, then $[c, b_n]$ does. Then choose

$$a_{n+1} = c \quad \text{and} \quad b_{n+1} = b_n\,.$$

As before,

$$b_{n+1} - a_{n+1} = \frac{1}{2}(b_n - a_n) \leq \frac{1}{2}\frac{1}{2^{n-1}}(b_1 - a_1) = \frac{1}{2^n}(b_1 - a_1)\,.$$

The intervals are constructed as desired; the intervals are zooming in on the progressively denser and denser parts of the sequence:



Now we apply the *Nested Intervals Theorem* to conclude that

$$a_n \to d \leftarrow b_n\,.$$

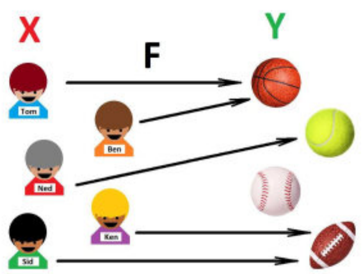
The second part of the construction is to choose the terms of the subsequence y_k of x_n, as follows. We just pick as y_k any element of the set $\{x_n : \ n = 1, 2, 3, ...\}$ in $[a_k, b_k]$ that comes later in the sequence than the ones already added, i.e., $y_1, y_2, ..., y_{k-1}$. This is always possible because we always have infinitely many elements left to choose from. Once the subsequence y_k is constructed, we have $y_k \to d$ by the *Squeeze Theorem*.

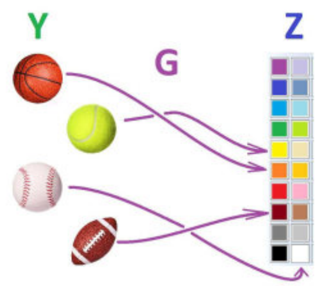
1.8. Compositions

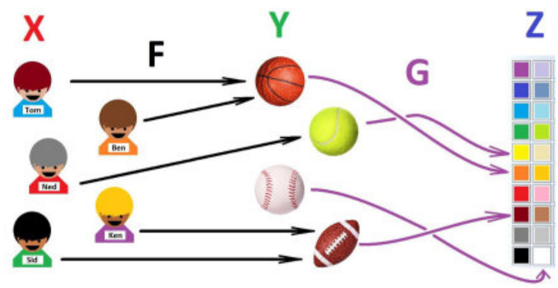
After all of these algebraic operations, what about the compositions?

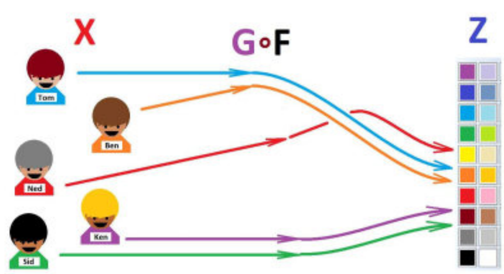
Let's review.

Suppose we have a set of boys and a set of balls and we know the boys' preferences in balls. That's one function. We also note the colors of the balls. That's another function. Do we know their *preferences in colors*? It's the composition of the two functions:









If we represent the two functions as *black boxes*, we can wire them together:

$$\begin{array}{ccccccccc}
\text{input} & & \text{function} & & \text{output} & & & & \\
x & \to & \boxed{F} & \to & y & & & & \\
 & & & & \downarrow & & & & \\
 & & & & \text{input} & & \text{function} & & \text{output} \\
 & & & & y & \to & \boxed{G} & \to & z
\end{array}$$

Thus, we use the output of the former as the input of the latter.

The idea is that the two (or more) functions are applied consecutively:

Definition 1.8.1: composition of functions

Suppose we have two functions (with the codomain of the former matching the domain of the latter):

$$F : X \to Y \quad \text{and} \quad G : Y \to Z\,.$$

Then their *composition* is the function (from the domain of the former to the codomain of the latter)

$$H : X \to Z\,,$$

which is computed for every x in X according to the following two-step procedure:

$$x \ \to \ F(x) = y \ \to \ G(y) = z\,.$$

Composition of functions

$$\underset{\text{name of the new function}}{\underset{\uparrow}{\big(\, G \circ F \,\big)}}\ (x) \quad = \quad \overset{\text{names of the second and first functions}}{\overset{\downarrow \qquad \downarrow}{G\big(\ \ F\ (x)\ \ \big)}}$$

$$\underset{\text{substitution}}{\uparrow \quad \uparrow \quad \uparrow}$$

Example 1.8.2: composition of numerical functions

This substitution is just as immediate for numerical functions. For example,

- $y = x^2$ is substituted into $z = y^3$, resulting in $z = \left(x^2\right)^3$.

The idea is the same as before (Volume 1, Chapter 1PC-3): Insert the input value in all of these boxes. For example, this function on the left is understood and evaluated via the diagram on the right:

$$f\,(y) = \frac{2y^2 - 3y + 7}{y^3 + 2y + 1}, \quad f\,(\square) = \frac{2\square^2 - 3\square + 7}{\square^3 + 2\square + 1}\,.$$

This is the result of the substitution $y = \sin x$:

$$f\boxed{(\sin x)} = \frac{2\boxed{(\sin x)}^2 - 3\boxed{(\sin x)} + 7}{\boxed{(\sin x)}^3 + 2\boxed{(\sin x)} + 1}\,.$$

Then, we have

$$f\,(\sin x) = \frac{2(\sin x)^2 - 3(\sin x) + 7}{(\sin x)^3 + 2(\sin x) + 1}\,.$$

Example 1.8.3: composition with sequences

The input of a sequence – understood as a function – is an integer, but the output could be any real number. Then, can we have the composition of two sequences:

$$\begin{array}{lllllll}
\text{integer input, } n & \to & \boxed{\text{sequence}} & \to & \text{real output, } a_n & & \\
 & & & & \downarrow ? & & \\
 & & & & \text{integer input, } m & \to & \boxed{\text{sequence}} \ \to \ \text{real output, } b_m
\end{array}$$

There is a mismatch of the variables! For example, what is the composition of these two sequences:

$$a_n = 1/n \ \text{ and } \ b_n = (-1)^n\,?!$$

However, if the first sequences is integer-valued, the composition makes sense.

$$\begin{array}{lllllll}
\text{integer input, } n & \to & \boxed{\text{sequence}} & \to & \text{integer output, } m_n & & \\
 & & & & \downarrow & & \\
 & & & & \text{integer input, } m & \to & \boxed{\text{sequence}} \ \to \ \text{real output, } b_m
\end{array}$$

For example, if we have:

$$a_n = 2n \ \text{ and } \ b_k = (-1)^k\,,$$

then their composition is

$$c_n = b_{2n} = (-1)^{2n} .$$

It's a subsequence of b_n!

If the second one is just a function of real variable, there is a composition. For example, the composition of

$$x_n = 1/n \quad \text{and} \quad f(x) = x^2$$

is

$$y_n = (1/n)^2 .$$

This is the flowchart of such a composition:

$$\begin{array}{llllllll}
\text{integer input, } n & \to & \boxed{\text{sequence}} & \to & \text{real output, } x_n & & & \\
 & & & & \downarrow & & & \\
 & & & & \text{real input, } x & \to & \boxed{\text{function}} & \to \text{real output, } y
\end{array}$$

Exercise 1.8.4

State as an implication: "Every sequence is a function." What about the converse?

For a function f defined on an interval, the following sequence is likely to make sense:

$$\boxed{y_n = f(x_n)}$$

Now, what about the limit of the new sequence? Can we say, similar to the four rules of limits, the limit of the *composition* is the composition of the limits?

The question is:

- ▶ What does the composition with a function do to the limit of a convergent sequence?

Let's look at some examples.

Example 1.8.5: linear

Sometimes the algebra is simple. Suppose we have a sequence $x_n \to a$. Suppose f is a linear polynomial:

$$f(x) = mx + b .$$

What is the limit of the composition sequence $y_n = f(x_n)$? Let's compute:

$$\begin{aligned}
\lim_{n\to\infty} y_n &= \lim_{n\to\infty} f(x_n) \\
&= \lim_{n\to\infty} (mx_n + b) \\
&= m \lim_{n\to\infty} x_n + b \qquad \text{According to the Linearity Rule.} \\
&= ma + b \\
&= f(a) .
\end{aligned}$$

This is just the value of f at a! In retrospect, the limit can be found by a simple substitution. For example,

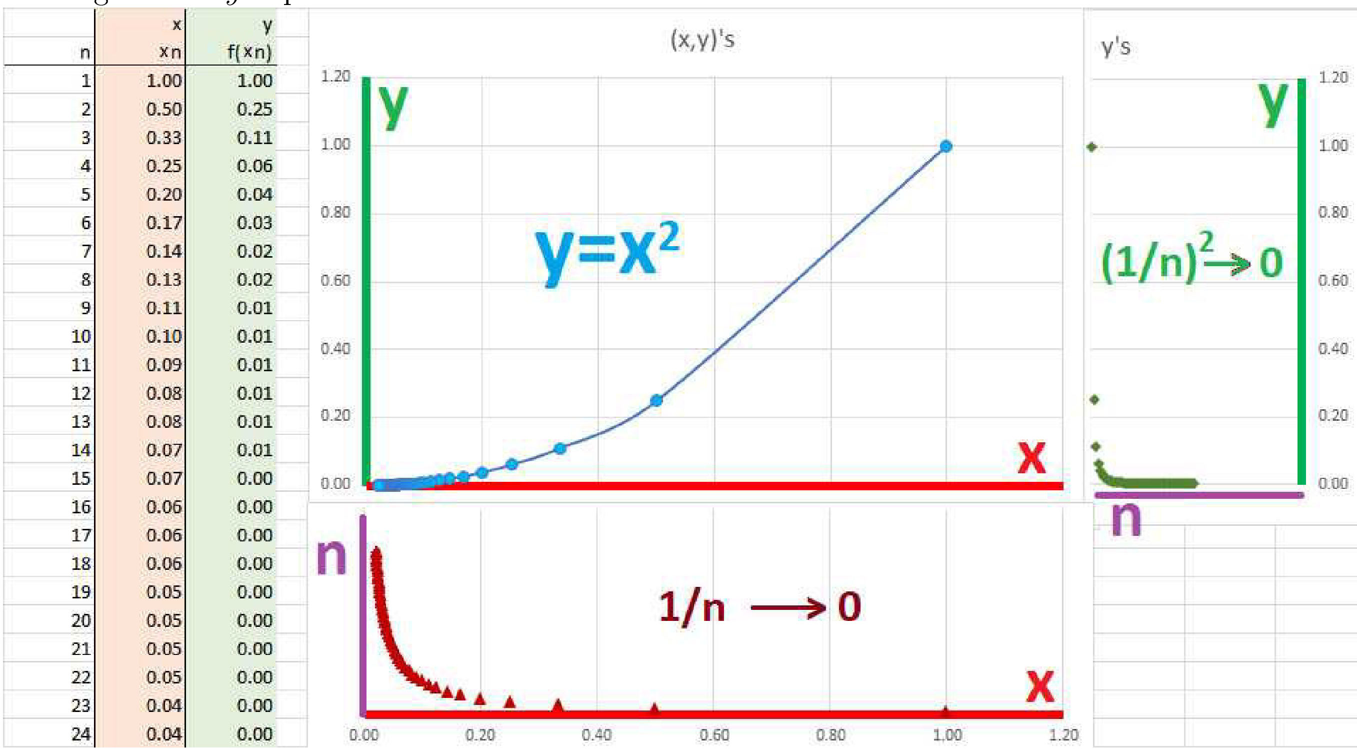
$$\lim_{n\to\infty} \left(3 \left(\frac{n+1}{n-1} \right) + 7 \right) = 3 \cdot \lim_{n\to\infty} \left(\frac{n+1}{n-1} \right) + 7 = 3 \cdot 1 + 7 = 10 .$$

So, the limit of the composition is the *value* of the function at the limit!

Example 1.8.6: compositions

Let's try $f(x) = x^2$ and a sequence that converges to 0. This is what we see below:

- bottom: how x depends on n,
- middle: how y depends on x,
- right: how y depends on n.



n	x xn	y f(xn)
1	1.00	1.00
2	0.50	0.25
3	0.33	0.11
4	0.25	0.06
5	0.20	0.04
6	0.17	0.03
7	0.14	0.02
8	0.13	0.02
9	0.11	0.01
10	0.10	0.01
11	0.09	0.01
12	0.08	0.01
13	0.08	0.01
14	0.07	0.01
15	0.07	0.00
16	0.06	0.00
17	0.06	0.00
18	0.06	0.00
19	0.05	0.00
20	0.05	0.00
21	0.05	0.00
22	0.05	0.00
23	0.04	0.00
24	0.04	0.00

Can we prove the convergence that we see? An application of the *Product Rule* in this simple situation reveals:

$$\begin{aligned}\lim_{n\to\infty} (x_n)^2 &= \lim_{n\to\infty} (x_n \cdot x_n)\\ &= \lim_{n\to\infty} x_n \cdot \lim_{n\to\infty} x_n\\ &= \left(\lim_{n\to\infty} x_n\right)^2,\end{aligned}$$

provided that limit exists.

A repeated use of the *Product Rule* produces a more general formula: If a sequence x_n converges, then so does $(x_n)^p$ for any positive integer p, and

$$\lim_{n\to\infty} [(x_n)^p] = \left[\lim_{n\to\infty} x_n\right]^p .$$

Combined with the *Sum Rule* and the *Constant Multiple Rule* this proves the following:

Theorem 1.8.7: Composition Rule for Limits of Polynomials

If sequence x_n converges, then so does $f(x_n)$ for any polynomial f. Furthermore, we have:

$$\boxed{\lim_{n\to\infty} f(x_n) = f\left(\lim_{n\to\infty} x_n\right)}$$

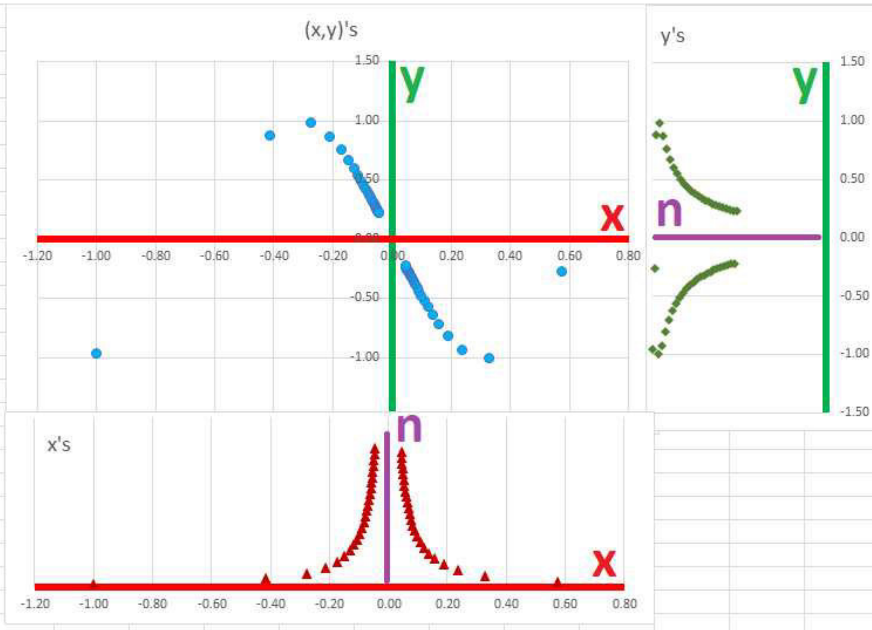
Such a result is also known as the *Substitution Rule* because the limit is substituted into this function as its input. It is a generalization of the *Linearity Rule*.

Example 1.8.8: from both sides

This time we choose a sequence that approaches 0 while alternating between the sides:

$$x_n = (-1)^n \frac{1}{n^{.8}} \quad \text{and} \quad f(x) = -\sin 5x\,.$$

	x	y
n	xn	f(xn)
1	-1.00	-0.96
2	0.57	-0.27
3	-0.42	0.87
4	0.33	-1.00
5	-0.28	0.98
6	0.24	-0.93
7	-0.21	0.87
8	0.19	-0.81
9	-0.17	0.76
10	0.16	-0.71
11	-0.15	0.67
12	0.14	-0.63
13	-0.13	0.60
14	0.12	-0.57
15	-0.11	0.54
16	0.11	-0.52
17	-0.10	0.50
18	0.10	-0.48
19	-0.09	0.46
20	0.09	-0.44
21	-0.09	0.42
22	0.08	-0.41
23	-0.08	0.40
24	0.08	-0.38
25	-0.08	0.37



We see the same pattern!

This is the summary of the theorem:

$$\boxed{x_n \to a \implies f(x_n) \to f(a)}$$

Proven only for polynomials!

Example 1.8.9: limits by substitution

It will be very easy to compute some limits now; for example:

$$\begin{aligned}
\lim_{n\to\infty} \left(3 \cdot \left(\frac{1}{n}\right)^2 - 2 \cdot \left(\frac{1}{n}\right) + 5 \right) &= \left(3 \cdot 0^2 - 2 \cdot 0 + 5\right) = 5 \\
\lim_{n\to\infty} \left(3 \cdot \left(2 + \frac{3}{n}\right)^2 - 2 \cdot \left(2 + \frac{3}{n}\right) + 5 \right) &= \left(3 \cdot 2^2 - 2 \cdot 2 + 5\right) = 13
\end{aligned}$$

We are using the same polynomial here:

$$f(x) = 3x^2 - 2x + 5\,.$$

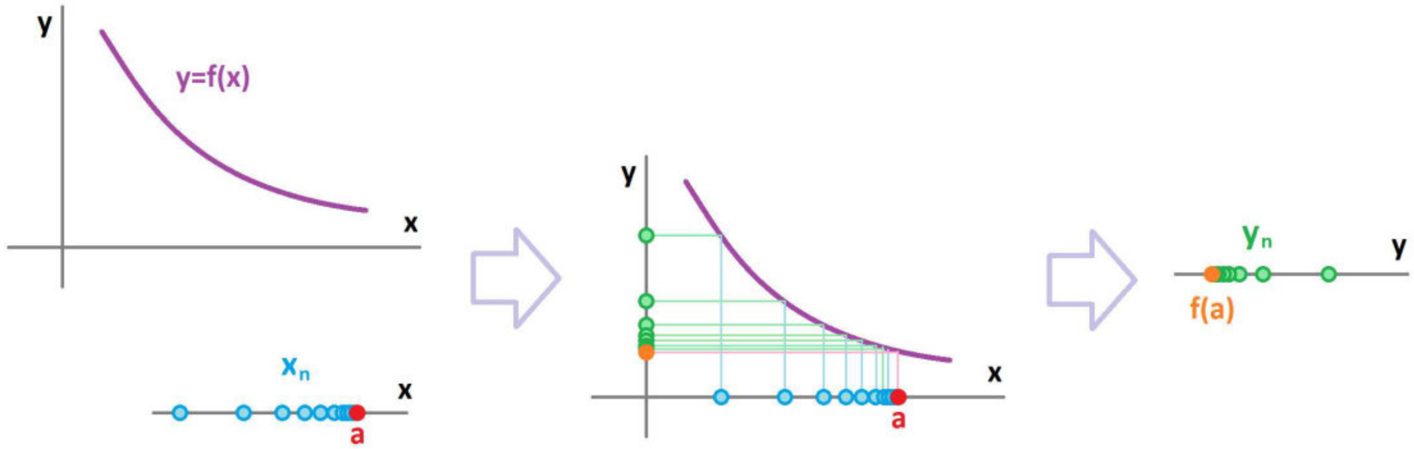
We simply substitute the appropriate limit, 0 and 2 respectively.

So, we conclude that limits behave well with respect to composition with *some* functions.

Once again, new sequences are produced via compositions with functions: Given a sequence x_n and a function $y = f(x)$, define

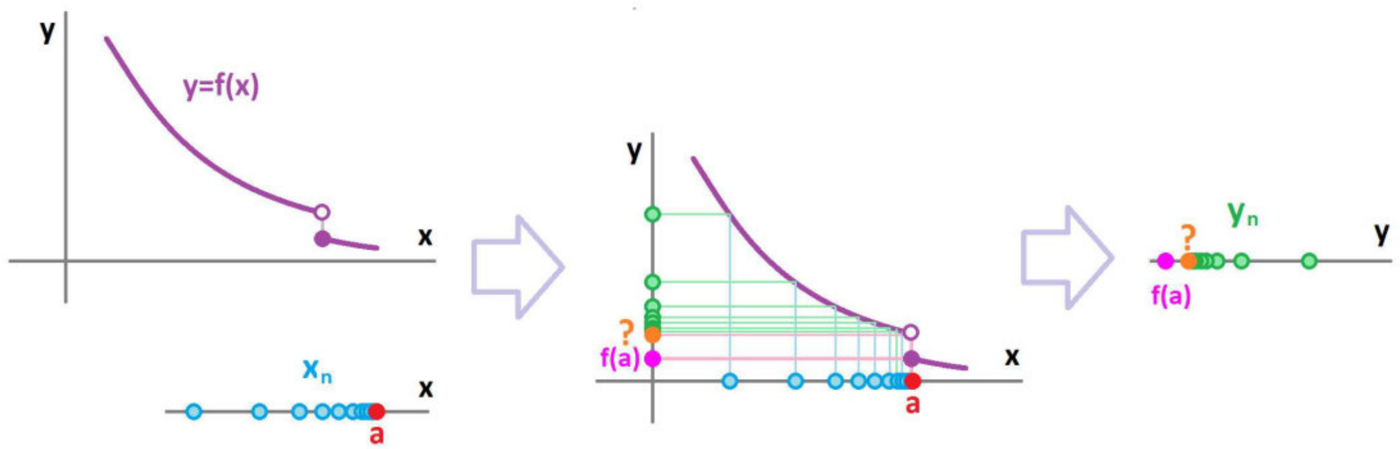
$$y_n = f(x_n)\,.$$

Let's examine the convergence of this sequence. Each point x_n on the x-axis produces a point y_n on the y-axis via f:



We can see how the convergence of x_n to a leads to the convergence of y_n to $f(a)$.

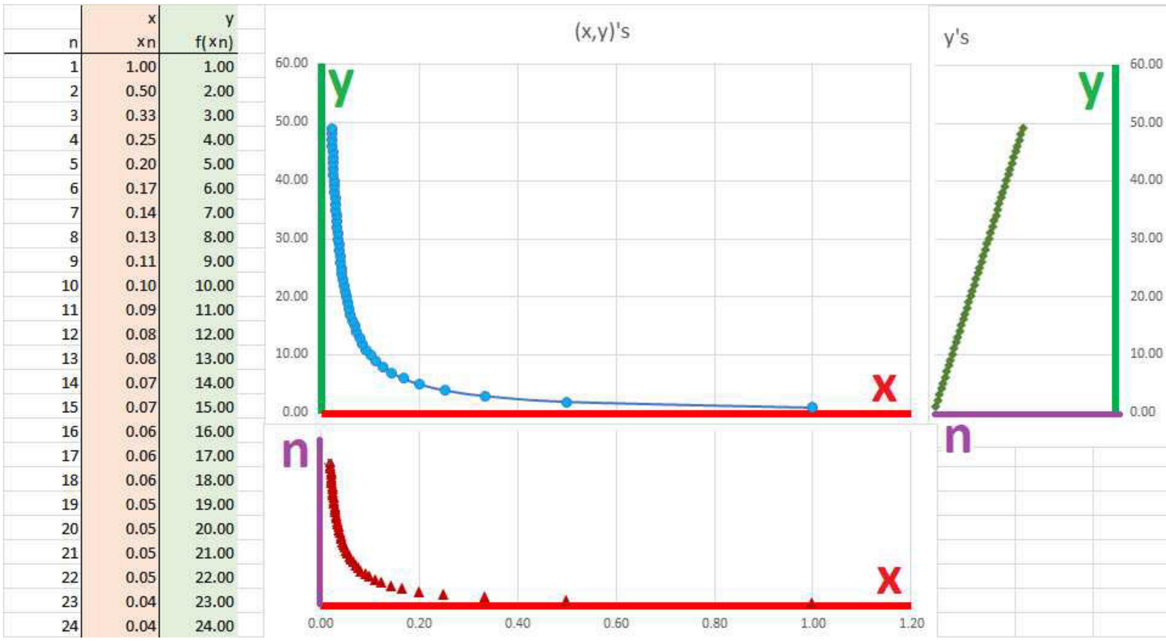
It is easy, however, to produce an example of a function for which this causation fails. Below, once again, each point x_n on the x-axis produces a point y_n on the y-axis via f:



However, because of the *gap* in the graph, y_n doesn't converge to $f(a)$! The issue is with *continuity vs. discontinuity* addressed in the next chapter.

Example 1.8.10: divergence to infinity

What if we choose
$$x_n = \frac{1}{n} \quad \text{and} \quad f(x) = \frac{1}{x}\,?$$
Then, obviously, we have:
$$y_n = \frac{1}{1/n} = n \to \infty\,!$$



n	x xn	y f(xn)
1	1.00	1.00
2	0.50	2.00
3	0.33	3.00
4	0.25	4.00
5	0.20	5.00
6	0.17	6.00
7	0.14	7.00
8	0.13	8.00
9	0.11	9.00
10	0.10	10.00
11	0.09	11.00
12	0.08	12.00
13	0.08	13.00
14	0.07	14.00
15	0.07	15.00
16	0.06	16.00
17	0.06	17.00
18	0.06	18.00
19	0.05	19.00
20	0.05	20.00
21	0.05	21.00
22	0.05	22.00
23	0.04	23.00
24	0.04	24.00

The *Composition Rule* fails for this function too!

In Chapter 2, we will use this construction to study the limits of *functions* rather than those of sequences. A few examples of that are presented in the two following sections: We will establish several important facts that will be used throughout the book.

1.9. Numbers are limits

So, limits (when finite) are numbers, and vice versa.

However, as we just saw, a number can be the limit of many sequences:

$$\begin{array}{cccccccc} .9 & .99 & .999 & .9999 & .99999 & \ldots & \to 1 \\ 1. & 1.1 & 1.01 & 1.001 & 1.0001 & \ldots & \to 1 \end{array}$$

Infinitely many, in fact:

$$\begin{array}{ccccc} a+1/n & & a+1/(2n) & & a+1/(3n) \\ & \searrow & \downarrow & \swarrow & \\ a+1/n^2 & \to & \boxed{a} & \leftarrow & a+1/n^3 \\ & \nearrow & \uparrow & \nwarrow & \\ a+1/\sqrt{n} & & a+1/2^n & & a+1/(\ln n) \end{array}$$

We, therefore, act in reverse:

▶ Instead of looking for the limit of a sequence, we find sequences that converge to a number we are interested in.

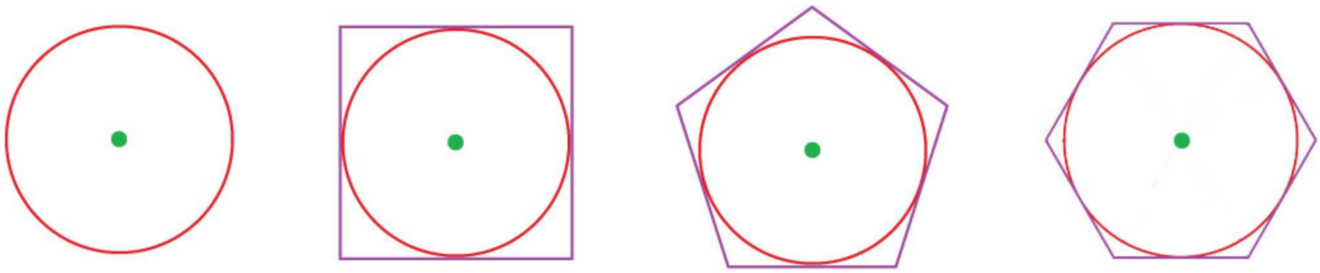
We think of those as *approximations* of this number.

Example 1.9.1: π

Let's build the next sequence from geometry:

▶ We approximate the circle with regular polygons: equal sides and angles.

We put such polygons around the circle so that it touches them from the inside ("circumscribing" polygons):



For each $n = 3, 4, 5, ...$, we find the area of the polygon. Let's call it A_n. We can examine the data:

n	3	4	5	6	7	8	9	10	11	12	13	14	15
A_n	5.196	4.000	3.633	3.464	3.371	3.314	3.276	3.249	3.230	3.215	3.204	3.195	3.188

Is there a *limit*? First, the sequence is bounded from below by 0. Second, it appears to be monotone. Then it must be convergent by the *Monotone Convergence Theorem*. We call its limit π! We can also put such polygons around the circle so that it touches them from the *outside* ("inscribing" polygons). These areas B_n also converge to π. We now have:

$$A_n \to \pi \leftarrow B_n\,.$$

Let's recall (Volume 1, Chapter 1PC-1) how real numbers are introduced.

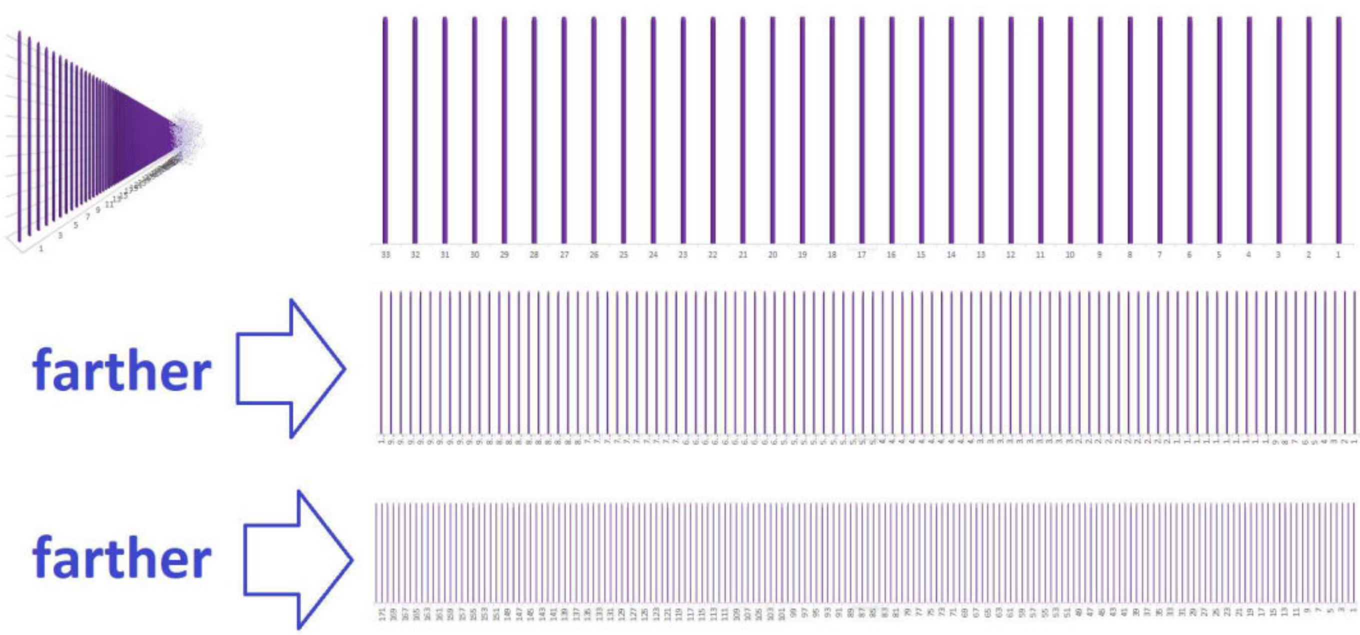
The starting point is the *natural numbers* (used for counting):

$$0, 1, 2, 3, ...$$

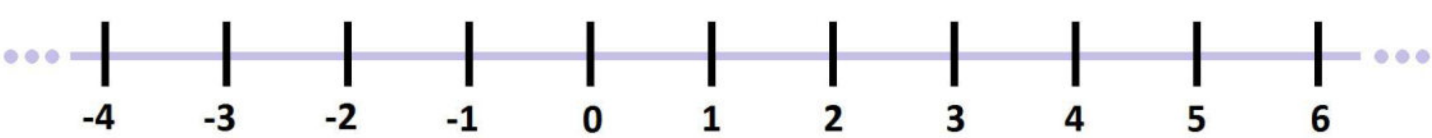
The next step is the *integers*:

$$..., -3, -2, -1, 0, 1, 2, 3, ...$$

They can be used for keeping record of space and locations. Imagine facing a fence so long that you can't see where it ends. We step *away* from the fence multiple times and there is still more to see:

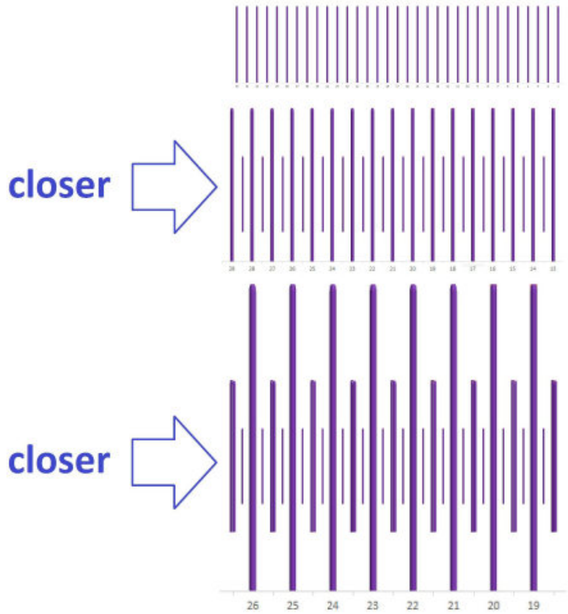


Is the number of planks *infinite*? We assume that it is. We visualize these as markings on a straight line, according to the order of the planks:



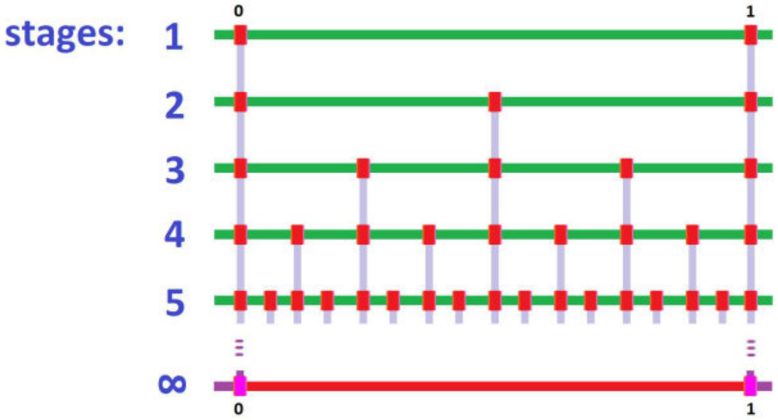
We have a sequence with an infinite limit!

So, we zoomed *out* to see the fence. Suppose now we zoom *in* on a spot on the fence. What if there is a shorter plank between every two planks? We look closer and we see more:



If we keep zooming in, the result will look similar to a *ruler*:

It's as if we add *one mark* between any two and then add another one between either of the two pairs we have created. We keep repeating this step. Ignoring the fact that this ruler goes only to 1/16 of an inch, let's imagine that the process continues indefinitely:

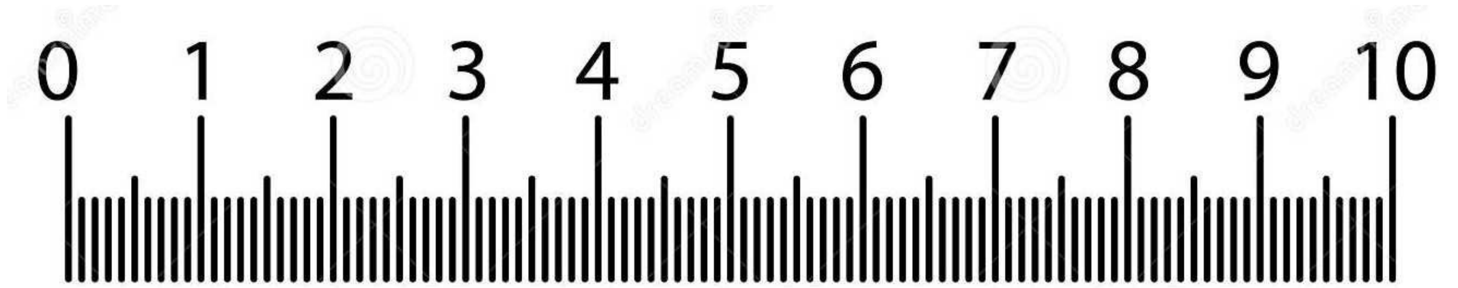


Is the depth *infinite*? We assume that it is. If we pick a point from each rung, we have bounded sequence.

Exercise 1.9.2

When does such a sequence converge?

If we add *nine marks* at a time, the result is a *metric ruler*:



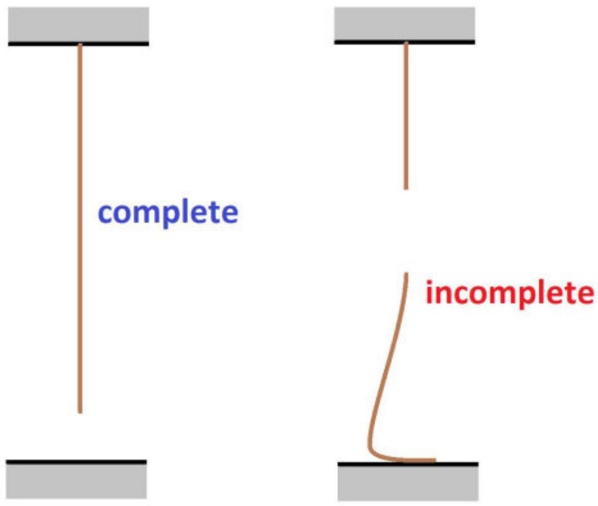
To see it another way, we allow more and more decimals in our numbers:

1.55 :	1.	1.5	1.55	1.550	1.5500	...
1/3 :	.3	.33	.333	.3333	.33333	...
1 :	1.	1.0	1.00	1.000	1.0000	...
π :	3.	3.1	3.14	3.141	3.1415	...

These are all convergent sequences!

So, we start with integers as locations and then – by cutting these intervals further and further – also include fractions, i.e., *rational numbers*.

However, we then realize that some of the locations have no counterparts among these numbers. For example, $\sqrt{2}$ is the length of the diagonal of a 1×1 square (and a solution of the equation $x^2 = 2$); it's not rational. Without this number, the line is incomplete! As an illustration, an "incomplete" rope won't hang:



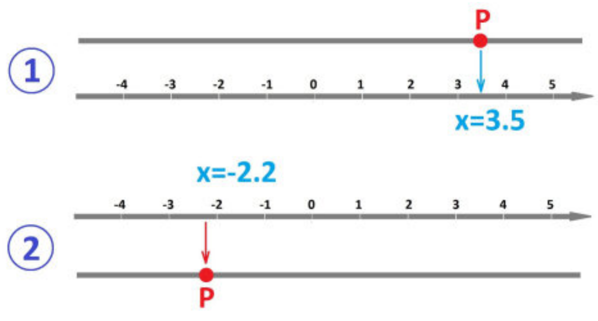
What to do? We approximate every such number with a sequence of rational numbers.

Together, rational and irrational numbers make up the *real numbers* and the *real number line*. We think of this line as *complete*; there are no missing points. The *Completeness Property of Real Numbers* proven above guarantees this.

This setup produces a correspondence between the locations on the line and the real numbers:

$$\boxed{\text{location } P \longleftrightarrow \text{number } x}$$

We will follow this correspondence in *both directions*, as follows:



- First, suppose P is a *location* on the line. We then find the corresponding mark on the line. That's the "coordinate" of P: some *number* x.
- Conversely, suppose x is a *number*. We think of it as a "coordinate" and find its mark on the line. That's the *location* of x: some point P on the line.

Once the coordinate system is in place, it is acceptable to think of every location as a number, and vice versa. In fact, we often write:

$$P = x\,.$$

The result may be described as the "1-dimensional coordinate system". It is also called the *real number line* or simply *the number line.*

But what about the algebra we routinely do with these numbers?

Our rules of limits show that when we replace every real number with a sequence converging to it, it is still possible to do algebraic operations with these replacements. Below is the summary:

Theorem 1.9.3: Algebra of Limits of Sequences

Suppose $a_n \to a$ and $b_n \to b$. Then

SR:	$a_n + b_n \to a + b$	*CMR:*	$c \cdot a_n \to ca$	*for any real c*
PR:	$a_n \cdot b_n \to ab$	*QR:*	$a_n / b_n \to a/b$	*provided $b \neq 0$*

Example 1.9.4: approximations

These rules show why *approximations work*. Indeed, we can think of a sequence that converges to a number as a sequence of better and better approximations. Then carrying out all the algebra with these sequences will produce the same result as the original computation is meant to produce! For example, here is such a substitution:

$$\begin{array}{ccccc} (1) & + & (2) & & = & 3 \\ \left(1+\dfrac{1}{n}\right) & + & \left(2-\dfrac{5}{n}\right) & = 3-\dfrac{4}{n} & \to & 3 \\ \left(1-\dfrac{3}{n}\right) & + & \left(2+\dfrac{2}{n^2}\right) & & \to & 3 \\ ... & & & & & \end{array}$$

Similarly, if we need to find $\sqrt{2}+\pi$, we use two approximations of the two numbers (no matter which ones):

$$a_n \to \sqrt{2} \quad\text{and}\quad b_n \to \pi\,.$$

Then the new sequence gives us an approximation of the sum:

$$a_n + b_n \to \sqrt{2} + \pi\,.$$

Example 1.9.5: decimals

These laws help us justify the following *trick* of finding fraction representations of infinite decimals.

This is how we deal with $x = .3333...$:

$$\begin{array}{ll} x & = 0.3333... \\ - & \\ 10x & = 3.3333... \\ \hline -9x & = -3.0000... \quad \text{Solve this equation!} \\ & \qquad\qquad\qquad \implies x = 1/3 \end{array}$$

Behind this trick is a proper method.

First, we note that according to the Monotone Convergence Theorem, the following sequence converges:

$$0 \quad 0.3 \quad 0.33 \quad 0.333 \quad 0.3333 \quad ... \quad \to \quad x$$

We call the limit x.

Then we use the *Constant Multiple Rule* and the *Difference Rule* to carry out the following algebra of sequences:

$$\begin{array}{lllllllll} a_n: & 0 & 0.3 & 0.33 & 0.333 & 0.3333 & ... & \to & x \\ - & & & & & & & & \\ 10a_n: & 3 & 3.3 & 3.33 & 3.333 & 3.3333 & ... & \to & 10x \\ \hline -9a_n: & -3 & -3 & -3 & -3 & -3 & ... & \to & -9x \\ & & & & & & & & \text{Solve: } -3 = -9x \implies x = 1/3 \end{array}$$

Note that we have shifted the values of the second sequence.

As we know, computing a polynomial only requires the first three algebraic operations. That is why we can approximate outputs of a polynomial f from approximations of the inputs:

$$a_n \to a \implies f(a_n) \to f(a)\,.$$

This follows from *Algebra of Limits of Sequences* and is just another version of *Composition Rule for Limits of Polynomials*. According to the *Quotient Rule*, we have this property for rational functions too as long as we avoid dividing by 0. But what about more complex functions?

1.10. The exponential functions

Let's recall first what we know about the algebra of exponents (seen in Volume 1, Chapter 1PC-3).

It starts with this simple algebra:

- Repeated addition is *multiplication*: $2 + 2 + 2 = 2 \cdot 3\,.$

One can say that that's how multiplication was "invented" – as repeated addition. Next:

- Repeated multiplication is *power*: $2 \cdot 2 \cdot 2 = 2^3\,.$

And this is the notation that we use:

Base and exponent

$$\underset{\substack{\uparrow\\ \text{base}}}{b}^{\overset{\text{exponent}}{\overset{\downarrow}{x}}}$$

So, this notation is nothing but a *convention.*

From the idea of the exponent as a repeated multiplication, we derive the following rules (for arbitrary $a, b > 0$):

Analogy: repeated addition vs. repeated multiplication

	Multiplication:	**Exponentiation:**
$x = 1, 2, ...$	$\underbrace{a + a + a + ... + a}_{x \text{ times}} = a \cdot x$	$\underbrace{a \cdot a \cdot a \cdot ... \cdot a}_{x \text{ times}} = a^x$
$x = 0$	$a \cdot 0 = 0$	$a^0 = 1$
$x = -1, -2, ...$	$ax = (-a)(-x)$	$a^x = \left(\frac{1}{a}\right)^{-x} = \frac{1}{a^{-x}}$
Rules: 1.	$a(x + y) = ax + ay$	$a^{x+y} = a^x a^y$
2.	$(a + b)x = ax + bx$	$(ab)^x = a^x b^x$
3.	$a(xy) = (ax)y$	$a^{xy} = (a^x)^y$

Initially, the *exponential function* of base $a > 0$ is defined to be the geometric progression:

$$\boxed{f(x) = b^x}$$

with the *domain* the set of all integers:

$$\mathbf{Z} = \{..., -3, -2, -1, 0, 1, 2, 3, ...\}\,.$$

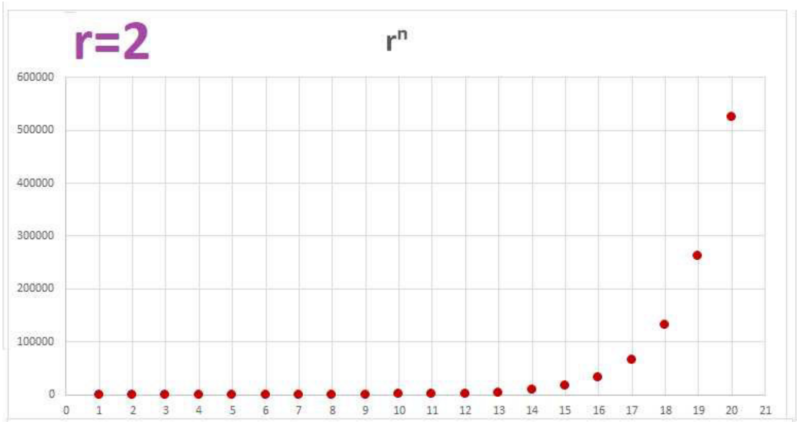
However, the domain misses some of the numbers that interest us!

Example 1.10.1: bacteria multiplying

Suppose we have a population of bacteria that doubles every day:

$$\underbrace{p_{n+1}}_{\text{population: at time } n+1} = 2 \cdot \underbrace{p_n}_{\text{at time } n} \implies p_n = p_0 2^n\,.$$

The graph consists of disconnected points:



Let's think of it as a function. It is given by the same formula, with x's still limited to the integers:

$$p(x) = p_0 2^x \,.$$

Now, what is the population in the middle of the first day? To answer, we consider the fact that multiplying by 2 is equivalent to multiplying $\sqrt{2}$ twice:

$$\sqrt{2} \cdot \sqrt{2} = 2 \,.$$

We conclude that

$$p(1/2) = \sqrt{2} \,.$$

Example 1.10.2: compounded interest

Consider what happens to a \$1000 deposit with 10% annual interest, compounded yearly. It's a geometric progression; after x years we have:

$$f(x) = 1000 \cdot 1.1^x \,.$$

where x is a positive integer.

But what if I want to withdraw my money in the *middle* of the year? It would be fair to ask the bank for the interest to be compounded now. It would also be fair for the bank to do it in such a way that the annual return remains the same even if we compound twice. What should be the semi-annual interest rate?

Suppose the amount is to grow by a proportion, r. Then, if applied again, it will give me the same ten percent growth! In other words, we have:

$$f(.5) = 1000 \cdot r \quad \text{and} \quad r \cdot r = 1.1 \,.$$

Therefore, according to the definition of the square root, we have

$$r = \sqrt{1.1} \approx 1.0488 \,,$$

or about 4.9 percent.

Exercise 1.10.3

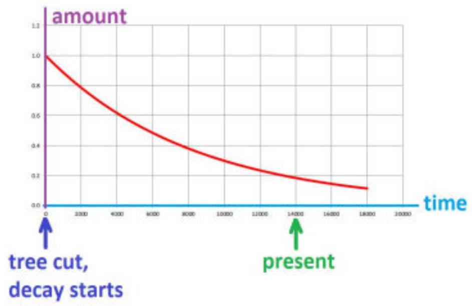
If your bank promises to pay you 1% for the first year and 2% for the second, what *single* (annual) interest can it offer in order to pay you as much over the next two years? Explain the meaning of the *average* interest rate.

Example 1.10.4: radioactive decay and radiocarbon dating

The radioactive carbon loses half of its mass over a certain period of time called the *half-life* of the element. It's a geometric progression again:

$$a_{n+1} = a_n \cdot \frac{1}{2}\,.$$

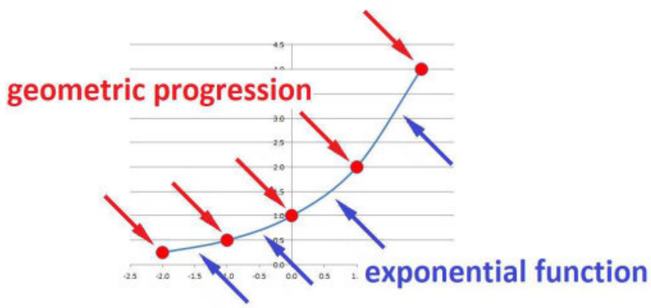
Unfortunately, n is not the number of years but the number of half-lives! For example, the percentage of this element, ^{14}C, left is plotted below against time:



However, we only know *two* points on the graph! Suppose the half-life is 5730 years (i.e., the time it takes to go from 100% to 50%). The model measures time in multiples of the half-life, 5730 years, and any period shorter than that will require a new insight. How much is left after $5730/2 = 2865$ years? The answer is below 75%:

$$\sqrt{\frac{1}{2}} \approx .707\,.$$

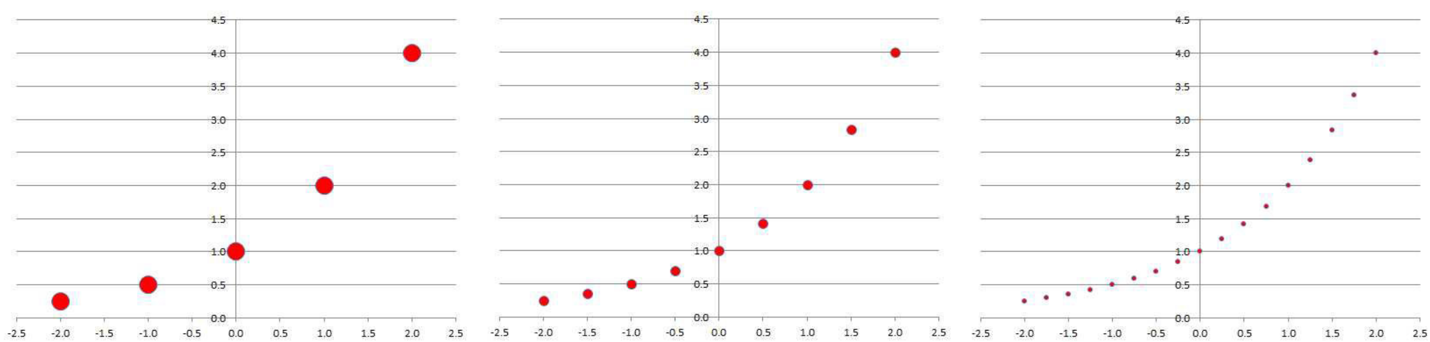
We fill the gaps in this manner:



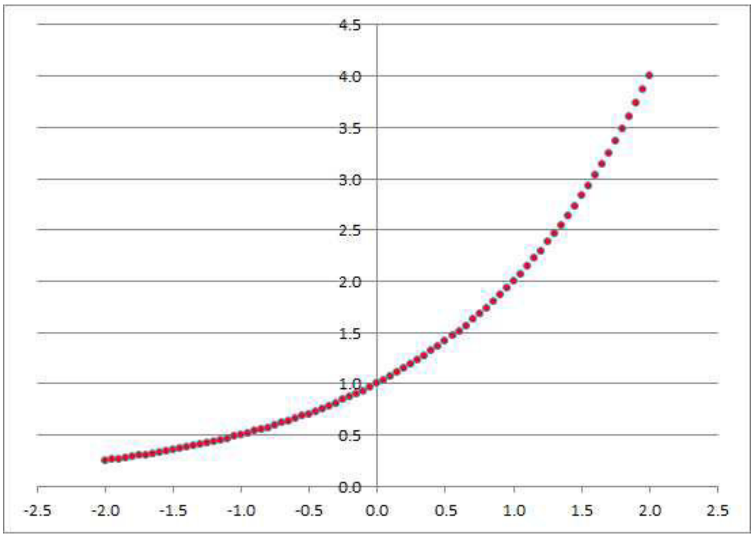
For any two numbers a and b, we know the value of the exponent for the number half-way between them. For example, we have the following:

$$x = \frac{a+b}{2} \implies 2^x = \sqrt{2^a \cdot 2^b}\,.$$

What if we continue to produce more and more values of this function by dividing the intervals in half? The new value will always lie slightly below the line that connects the two points on the graph:



One can also imagine that the x-axis, as the domain, is becoming more and more densely covered. These initially loose points start to form a curve:



The curve will lie below any chord that connects two points on the graph. Such a function is called "concave up" (Chapter 3).

More generally, if $x = \dfrac{m}{n}$, where m, n are integers with $n > 0$, then we set the *xth power* of $b > 0$ to be the following:

$$\boxed{b^{\frac{m}{n}} = \sqrt[n]{b^m}}$$

Below is the summary of the rules of exponents that will reappear many times:

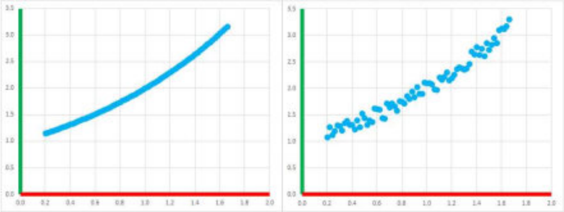
Theorem 1.10.5: Rules of Exponents

The identities below are satisfied for each rational $a > 0$ and every real x and y:

$$\begin{array}{lll} 1. & b^{x+y} & = b^x b^y \\ 2. & b^x c^x & = (bc)^x \\ 3. & b^{xy} & = (b^x)^y \end{array}$$

What does the graph of this new function look like? We have seen that, as these fractions get larger and larger denominators, the domain becomes denser and denser, so does the graph and, eventually, it becomes a curve! After all, the domain contains all rational numbers $\mathbf{Q}$ and the rational numbers are so dense that any interval, no matter how small, will contain infinitely many of them (indeed, if p and q are rational, then so is $(p+q)/2$). Therefore, there are *no gaps* in the graph!

However, its points remain *disconnected* from each other; a little blow and the curve falls apart:



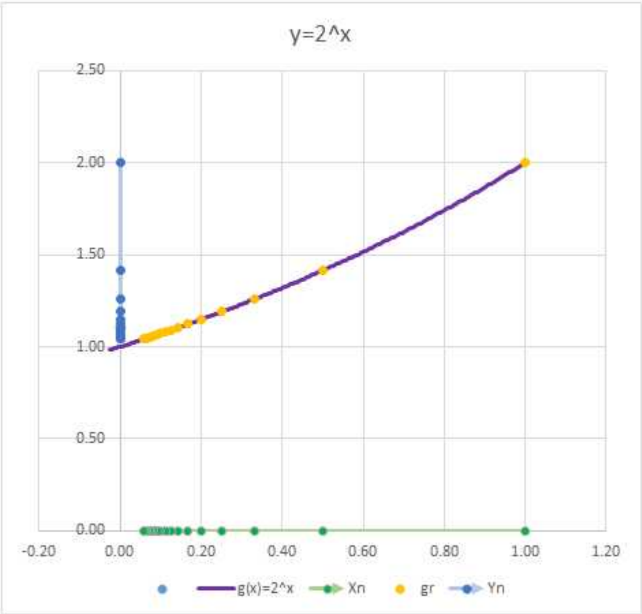
The reason is that the points on the graph with *irrational* x-coordinates are missing:

$$x = \sqrt{2},\ \sqrt{3},\ \pi\,.$$

And the irrational numbers are also so *dense* that any interval, no matter how small, will contain infinitely many of them. So, even though there are no gaps in the graph, there are invisible "cuts" everywhere!

We will make a step toward defining the function for the irrational exponents by "approximating" them with rational numbers.

The starting point will be to consider what is happening around $x = 0$. We can see a sequence on the x-axis that converges to 0 and the corresponding sequence on the y-axis converges to 1:



There are no jumps. In other words, we have:

Theorem 1.10.6: Exponential Function Around 0

For each real $b > 0$, we have:

$$\lim_{n\to\infty} b^{x_n} = 1$$

for every sequence of numbers $x_n \to 0$.

Proof.

Let's consider the simplified version of the statement:

$$\lim_{n\to\infty} a^{1/q_n} = 1\,,$$

where q_n is a sequence of integers with $q_n \to \infty$. To prove the formula, we need to find out for what values of n the inequality below is satisfied for any given $\varepsilon > 0$:

$$|\sqrt[q_n]{b} - 1| < \varepsilon\,.$$

For $b > 1$, the inequality is further simplified:

$$\sqrt[q_n]{b} - 1 < \varepsilon \iff \sqrt[q_n]{b} < 1 + \varepsilon \iff b < (1+\varepsilon)^{q_n}\,.$$

Now we can answer the question. We use this fact:

$$q_n \to \infty \implies (1+\varepsilon)^{q_n} \to \infty\,.$$

It follows that there is such an N that the last inequality is satisfied for all $n > N$. Then so is the original inequality.

Exercise 1.10.7

Provide the missing parts of the proof.

Just as we need only one quadratic polynomial, $y = x^2$, as a *template* for the rest, we really only need one exponential function!

Which one is the *one*?

Example 1.10.8: compounded interest

Suppose we have money in the bank at 10% APR compounded annually. Then after a year, given \$1,000 initial deposit, you have:

$$\begin{aligned} 1000 + 1000 \cdot 0.10 &= 1000(1 + 0.1) \\ &= 1000 \cdot 1.1\,. \end{aligned}$$

Same every year. After t years, it's $1000 \cdot 1.1^t$.

What if we want to compound more often with the goal of receiving the same interest at the end of the year? The exponential function is designed to solve this problem.

So, the account is compounded semi-annually, with the same APR as follows. After $\frac{1}{2}$ year, it's $1000 \cdot 0.05$, or a total of:

$$1000 + 1000 \cdot 0.05 = 1000 \cdot 1.05\,.$$

After another $\frac{1}{2}$ year, it is

$$(1000 \cdot 1.05) \cdot 1.05 = 1000 \cdot 1.05^2\,.$$

After t years,

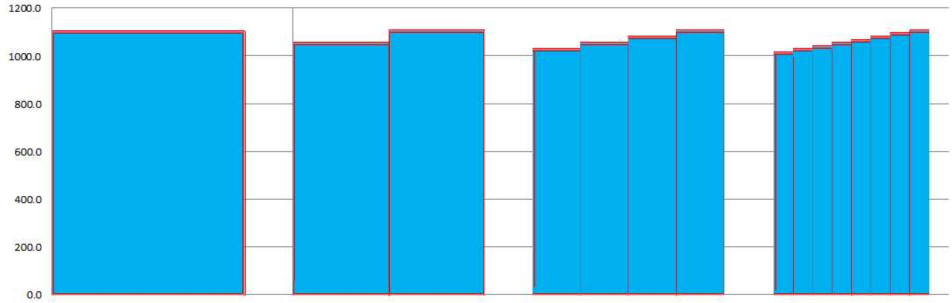
$$1000 \cdot (1.05^2)^t = 1000 \cdot 1.05^{2t}\,.$$

Note that we are getting more money than before: $1.05^2 = 1.1025 > 1.1$!

Next, we try to compound quarterly,

$$1000 \cdot 1.025^{4t}\,.$$

And so on. Below we see respectively: 5% compounded 1 time; $5/2 = 2.5\%$ compounded 2 times; $5/4 = 1.25\%$ compounded 4 times; $5/8 = .635\%$ compounded 8 times:



When the interest is compounded n times, we have:

$$1000 \cdot \left(1 + \frac{1}{n}\right)^{nt}\,,$$

where $\frac{1}{n}$ is the interest in one period. Is there a bound to this growth?

Exercise 1.10.9

Which of the four areas is the biggest?

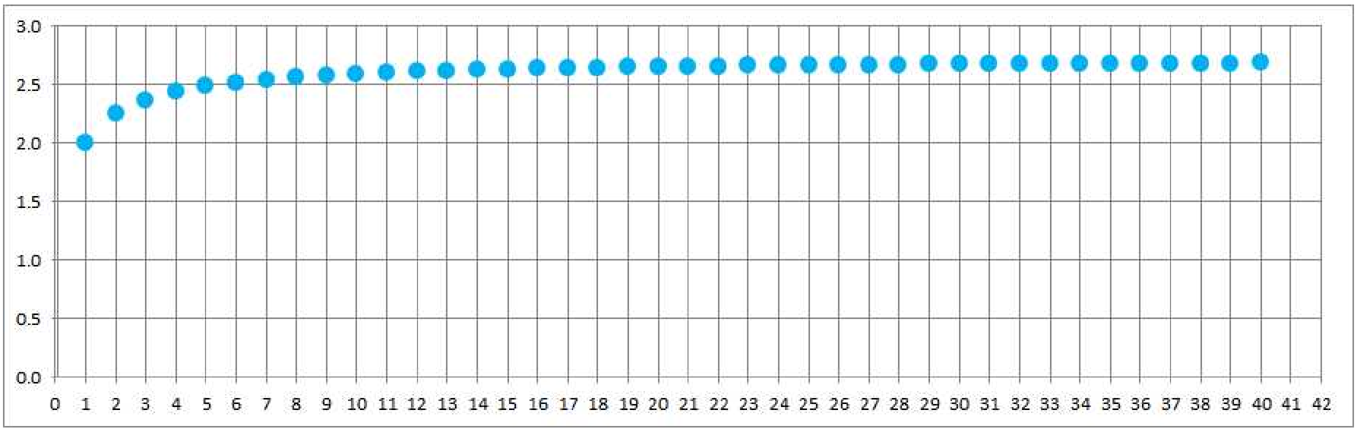
Generally, for APR r (given as a decimal) and for the initial deposit A_0, after t years, the current amount is

$$A(t) = A_0 \left(1 + \frac{r}{n}\right)^{nt}\,,$$

if compounded n times per year.

What if we compounded more and more frequently? Will we be paid unlimited amounts?

Let's pick a specific $r = 1$ (i.e., 100% APR) and plot this sequence. It appears that the growth slows down:



The answer to the question is No, and we prove this fact below:

Theorem 1.10.10: e as Limit

The limit below exists:

$$\boxed{\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n}$$

We have a special notation for this number:

e

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n$$

It is known as the "Euler number".

Proof.

First, we show that the sequence

$$a_n = \left(1+\frac{1}{n}\right)^n$$

is increasing. We have:

$$\begin{aligned}\frac{a_{n+1}}{a_n} &= \frac{\left(1+\frac{1}{n+1}\right)^{n+1}}{\left(1+\frac{1}{n}\right)^n} = \frac{\left(\frac{n+2}{n+1}\right)^{n+1}}{\left(\frac{n+1}{n}\right)^n} \\ &= \left(\frac{n+2}{n+1}\right)^{n+1}\left(\frac{n}{n+1}\right)^{n+1}\left(\frac{n+1}{n}\right)^1 \\ &= \left(\frac{n^2+2n+1-1}{n^2+2n+1}\right)^{n+1}\frac{n+1}{n} \\ &= \left(1-\frac{1}{(n+1)^2}\right)^{n+1}\frac{n+1}{n}\,.\end{aligned}$$

From the *Binomial Formula* (seen in Volume 1, Chapter 1PC-1), it follows that

$$(1+b)^m > 1+mb\,,$$

for any b. We just choose:

$$b = \frac{-1}{(n+1)^2} \quad\text{and}\quad m = n+1\,,$$

and then substitute the inequality into the first term of the last formula:

$$\begin{aligned}\frac{a_{n+1}}{a_n} &= (1+b)^m \cdot \frac{n+1}{n} \\ &> (1+mb)\cdot \frac{n+1}{n} \\ &= \left(1-\frac{1}{(n+1)^2}\right)^{n+1} \frac{n+1}{n} \\ &= \left(1-\frac{1}{n+1}\right)\frac{n+1}{n} \\ &= \frac{n}{n+1}\frac{n+1}{n} \\ &= 1\,.\end{aligned}$$

In a similar fashion, we show that the sequence

$$b_n = \left(1+\frac{1}{n}\right)^{n+1}$$

is decreasing. Since $a_n < b_n$, we conclude that former sequence is both increasing and bounded. Therefore, it converges by the *Monotone Convergence Theorem.*

The value of this special number is approximately the following:

$$e \approx 2.71828\,.$$

If you have 100% APR, this is the most you can get for your dollar no matter how frequently the interest is compounded. We have demonstrated the value of the following concept:

Definition 1.10.11: natural exponential function

The exponential function e^x with base e, the Euler number, is called the *natural base exponent.*

In order to use this function, we will need the following result that we will accept without proof:

Theorem 1.10.12: Exponential Function as Limit

$$\boxed{e^x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n}$$

Example 1.10.13: continuously compounding interest

We continue with the example: What if the interest is compounded n times and $n \to \infty$? Then we have:

$$\begin{aligned}\lim_{n\to\infty} A(x) &= \lim_{n\to\infty} A_0\left(1+\frac{r}{n}\right)^{nt} \\ &= A_0 \lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{nt} && \text{This is by CMR.} \\ &= A_0\left[\lim_{n\to\infty}\left(1+\frac{r}{n}\right)^{n}\right]^t && \text{This is by the theorem.} \\ &= A_0(e^r)^t\,.\end{aligned}$$

Thus, with an APR of r and an initial deposit A_0, after t years you have:

$$A(t) = A_0 e^{rt}\,.$$

We say that the interest is compounded *continuously*.

Suppose the APR is 10%, and $A_0 = 1000$, $x = 1$. Then:

$$A(1) = 1000 \cdot e^{1.1} = 1000 \cdot e^{0.1} \approx \$1,105\,.$$

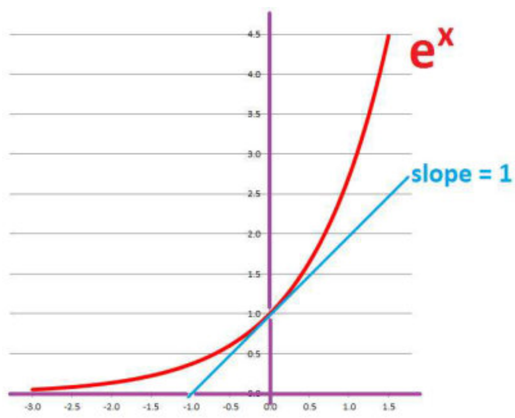
The interest is \$105 > \$100.

How long does it take to triple your money with the APR=5%, compounded continuously? Set $A_0 = 1$ and solve for t:

$$\begin{array}{rcll} 3 & = & 1 \cdot e^{0.05t} & \Longrightarrow \\ \ln 3 & = & 0.05t & \Longrightarrow \\ t & = & \dfrac{\ln 3}{0.05} & \approx 22 \text{ years.} \end{array}$$

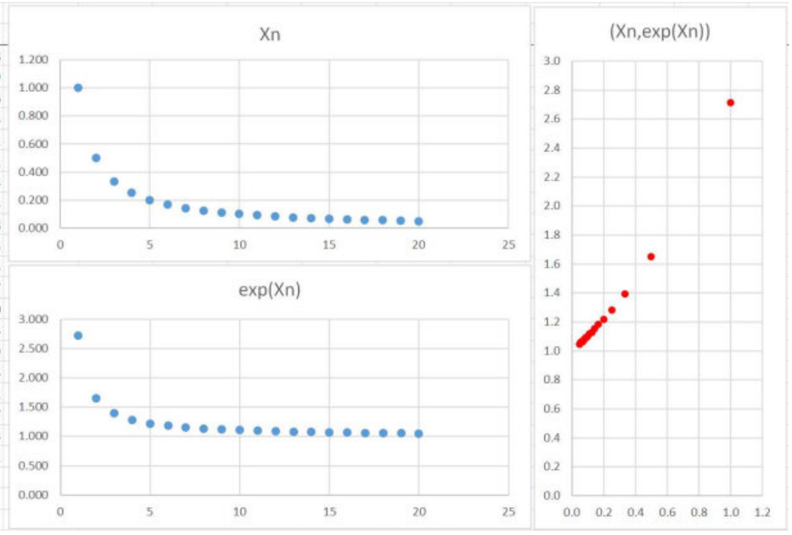
There is a more subtle observation about the graph of this function.

What makes e special is this property: The graph of $y = e^x$ almost merges with the line $y = x + 1$ around 0:



In fact, plotting the points $(1/n, e^{1/n})$ reveals a straight line with slope 1:

	1	
n	Xn	exp(Xn)
1	1.000	2.718
2	0.500	1.649
3	0.333	1.396
4	0.250	1.284
5	0.200	1.221
6	0.167	1.181
7	0.143	1.154
8	0.125	1.133
9	0.111	1.118
10	0.100	1.105
11	0.091	1.095
12	0.083	1.087
13	0.077	1.080
14	0.071	1.074
15	0.067	1.069
16	0.063	1.064
17	0.059	1.061
18	0.056	1.057
19	0.053	1.054
20	0.050	1.051



This is how we state this fact algebraically:

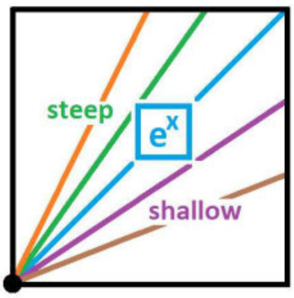
Theorem 1.10.14: Famous Limit: e^x

We have:

$$\lim_{n\to\infty} \frac{e^{x_n}-1}{x_n} = 1$$

for any sequence $x_n \to 0$.

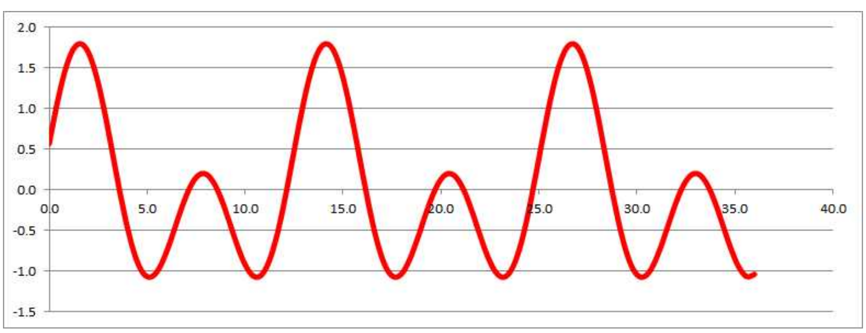
So, $y = e^x$ cuts, in a sense, the exponential functions into two halves: the "steep" ones and the "shallow" ones:



1.11. The trigonometric functions

Let's review what we know about these functions (Volume 1, Chapter 1PC-4).

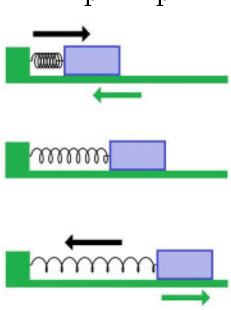
One encounters numerous examples of *periodic phenomena*. The simplest case is that of a quantity that changes but then comes back to change again in the same manner:

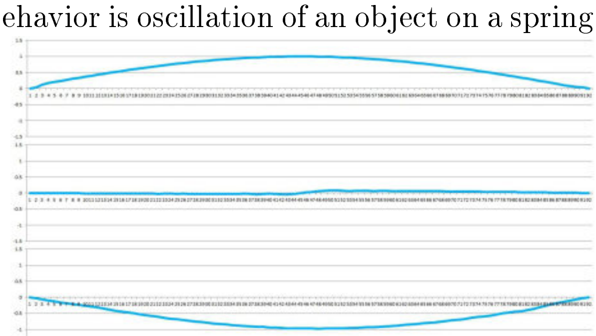


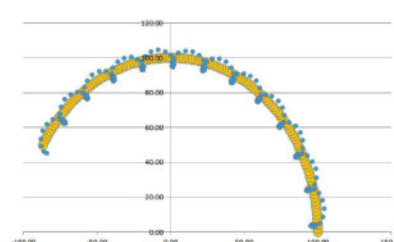
Functions with this behavior are called *periodic*.

Example 1.11.1: periodic behavior

The simplest periodic behavior is oscillation of an object on a spring or a string of a musical instrument:

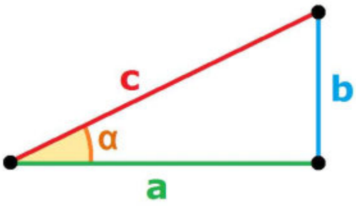






A more complex example is the trip of the moon around the sun.

The trigonometric functions initially come from *plane geometry*:



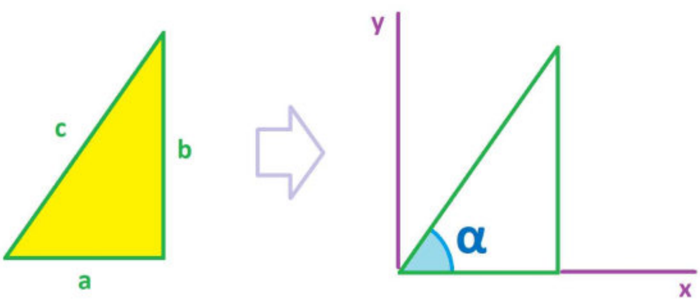
Suppose we have a right triangle with sides a, b, c, with c being the longest one facing the right angle. If α is the angle adjacent to side a, then we define the *cosine* and the *sine* of this angle as follows:

$$\begin{aligned}\cos\alpha &= \frac{a}{c}\\ \sin\alpha &= \frac{b}{c}\\ \tan\alpha &= \frac{b}{a} &= \frac{\sin\alpha}{\cos\alpha}\end{aligned}$$

The importance of the tangent is seen in this formula:

$$\tan\alpha = \frac{b}{a} = \frac{\text{rise}}{\text{run}} = \text{slope of the hypotenuse } c\,,$$

if side a follows the x-axis:



Slopes are further discussed in Chapter 3.

We adopt the **convention** that the size of the half-turn angle – and the length of the half-circle – is equal to π *radians*:

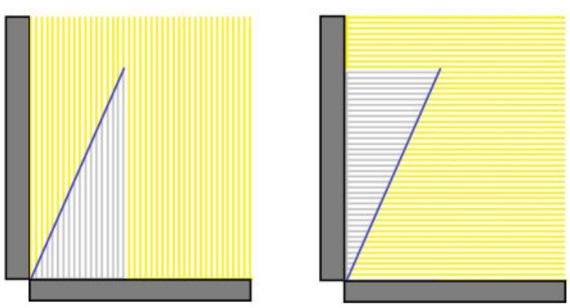
$$180 \text{ degrees} = \pi \text{ radians}$$

We, therefore, have the following conversion formulas for quantities measured in these units:

$$\#\text{ degrees} \quad = \#\text{ of radians} \cdot \frac{180}{\pi}$$

Example 1.11.2: shadow

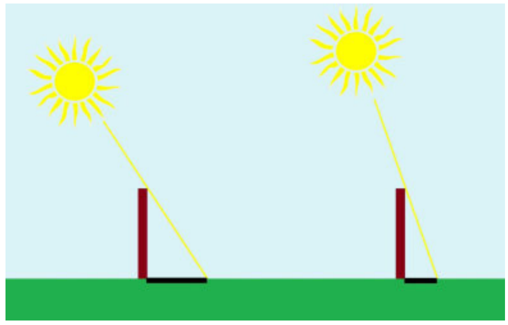
We have other interpretations of these quantities. Suppose we place a stick of length 1 at the angle α with the ground.



Then:

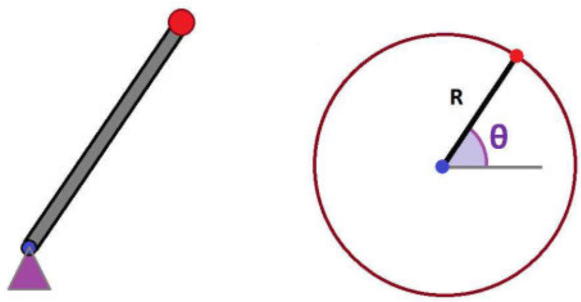
- We can think of $\cos\alpha$ as the length of its shadow – on the ground at noon (left).
- We can think of $\sin\alpha$ as the length of its shadow – on the wall at sunset (right).

Alternatively, the stick is vertical and still and it is the sun that is moving:



Example 1.11.3: rotating rod

Suppose a rod of length R is rotated around its end. If we can control the angle, θ, then what do we know about the *position* of its other end in space, i.e., its x and y coordinates?



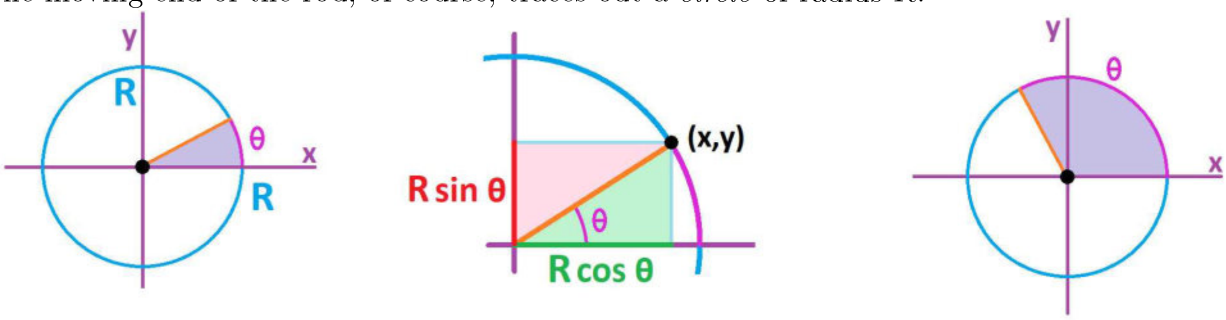
We use the above formulas:

$$\cos\theta = \frac{x}{R} \quad \text{and} \quad \sin\theta = \frac{y}{R}\,.$$

Therefore, our coordinates are given by these trigonometric functions:

$$x = R\cos\theta \quad \text{and} \quad y = R\sin\theta\,.$$

The moving end of the rod, of course, traces out a *circle* of radius R:



This analysis will be used as a basis for the polar coordinate system (Volume 3, Chapter 3IC-4).

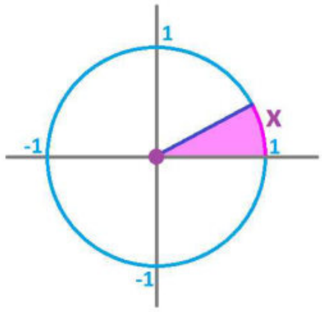
In a triangle, the value of angle α cannot go beyond 90 degrees and the formulas above give us 1/4 of the whole circle. We define the angle beyond 90 degrees:

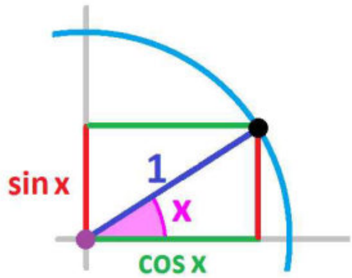
► The angle isn't an angle of a triangle but the angle of *rotation*.

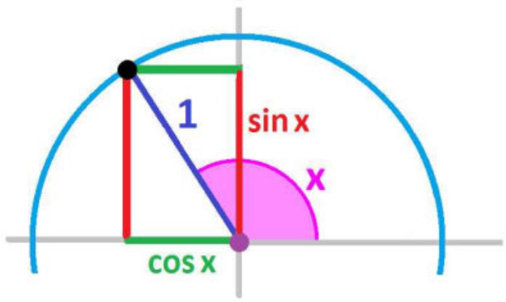
We will use the circle of radius 1.

Below, the name for the independent variable that represents the angle and runs within $(-\infty,+\infty)$ is x.

The outline of the construction is shown below:







We formalize this construction below:

Definition 1.11.4: sine and cosine functions

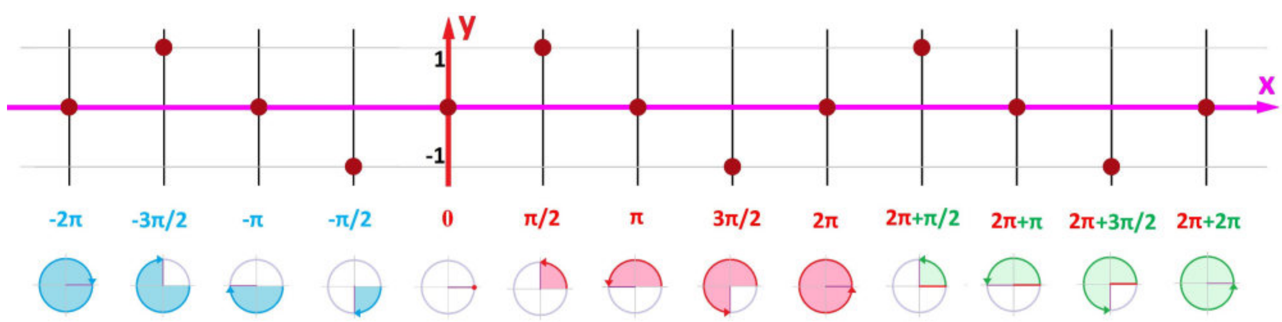
Suppose a real number x is given. We construct a line segment of length 1 on the Cartesian plane (with the horizontal axis *not* marked x) starting at 0 with angle x radians from the horizontal, counterclockwise. Then:

- The *cosine* of x is the horizontal coordinate of the end of the segment.
- The *sine* of x is the vertical coordinate of the end of the segment.

The positive direction of the angle, which is x, is *counterclockwise* (first row):

x	-2π	$-3\pi/2$	$-\pi$	$-\pi/2$	0	$\pi/2$	π	$3\pi/2$	2π	...
y	0	1	0	-1	0	1	0	-1	0	...

The values of the sine are provided in the second row. We plot the graph of $y = \sin x$ – point by point – following this table:



There is a similar table for the cosine.

How do we fill the gaps in the graph? We divide the intervals in half and apply the following trigonometric formula (seen in Volume 1, Chapter 1PC-5):

$$\sin\left(\frac{\alpha+\beta}{2}\right) = \pm\sqrt{\frac{1-\cos\alpha\cos\beta-\sin\alpha\sin\beta}{2}}$$

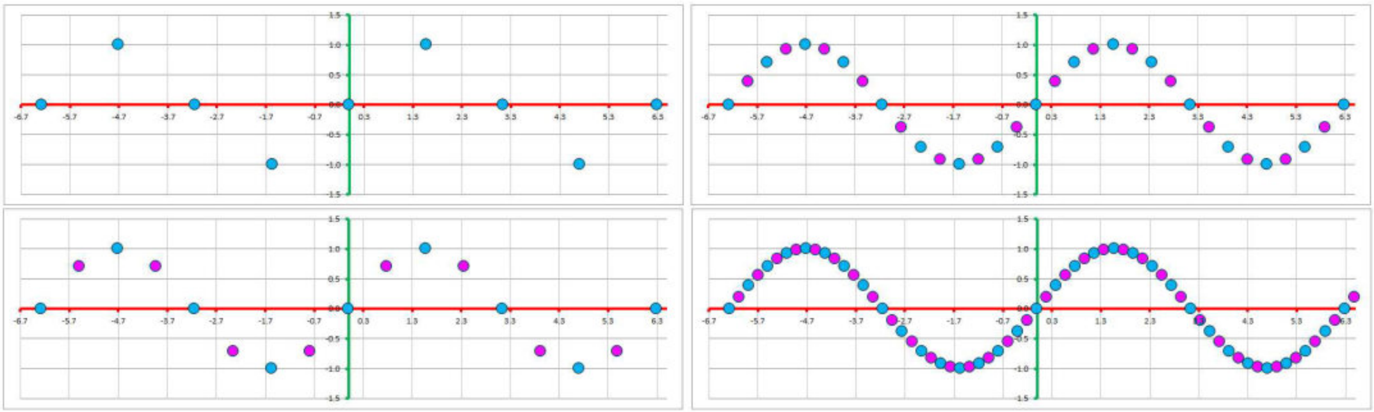
For example, this is how we defined the sine (and then cosine) of the angle halfway between 0 and $\pi/2$:

$$\sin\frac{\pi}{4} = \sin\left(\frac{0+\pi/2}{2}\right) = \sqrt{\frac{1-\cos 0\cos\pi/2-\sin 0\sin\pi/2}{2}} = \sqrt{\frac{1-\cos 0\cos\pi/2-\sin 0\sin\pi/2}{2}} = \sqrt{\frac{1}{2}}\,.$$

At the next stage, we then compute with the same formula the sine (and then cosine) of the angle half way between 0 and $\pi/4$:

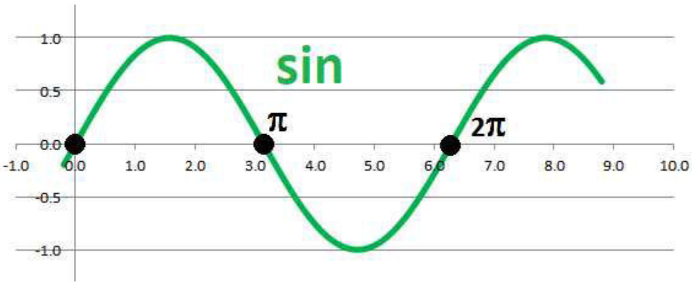
$$\sin\frac{\pi}{8} = \sin\left(\frac{0+\pi/4}{2}\right) = ...$$

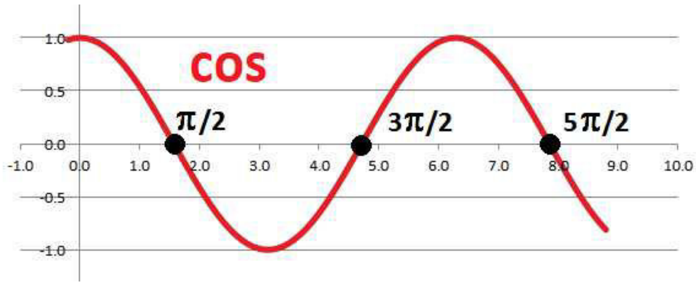
And so on. We continue to divide the intervals in half, producing more and more points on the graph:



The steps above, respectively, are: $\pi/2$, $\pi/4$, $\pi/8$, and $\pi/16$.

With more values found, these are the graphs of the sine and the cosine:





Even though these look like two curves, not all the locations have been filled!

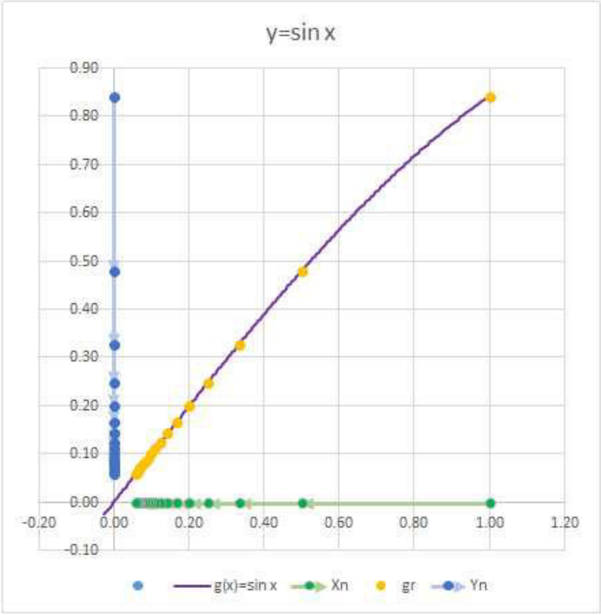
Indeed, the only values of x we have taken care of come from division by 2 starting with π. They are so close to each other that what is left is just invisible "cuts". But what is $\sin 1$?

We know that we need to approximate, and we are making the first step below.

We know that

$$\sin 0 = 0 \quad \text{and} \quad \cos 0 = 1\,.$$

Now, the two graphs seem to cross the y-axis at these two values of y. Let's make sure that there are *no jumps*. Just as before, we consider sequences on the x-axis that converge to 0:



What is happening to the corresponding y's? The picture suggests that they also converge to 0. And there is more:

Theorem 1.11.5: Sine and Cosine Around 0

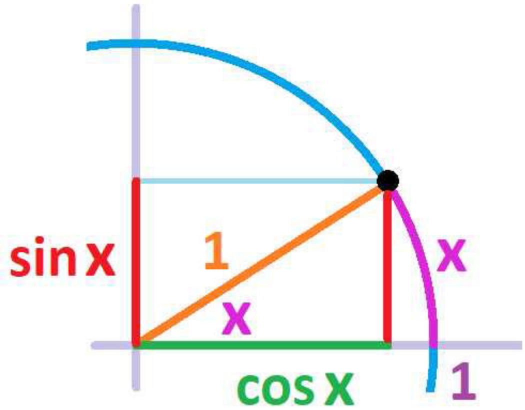
We have:

$$\boxed{\lim_{n\to\infty} \sin x_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \cos x_n = 1}$$

for any sequence $x_n \to 0$.

Proof.

The conclusion follows from certain facts from trigonometry (seen in Volume 1, Chapter 1PC-4):



We know the following about s segment of the circle of radius 1 with angle x:

- The length of the vertical segment is $\sin x$.
- The length of the curved segment is x.

Therefore, we have:

$$0 < \sin x < x\,.$$

If we have a sequence x_n that converges to 0, then so does the one in the middle by the *Squeeze Theorem*.

Warning!

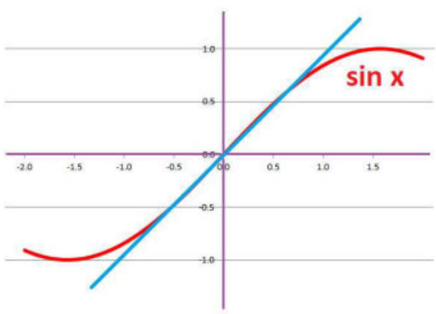
We will still need to confirm that our application of the idea of "length" to a curve is valid (Volume 3, Chapter 3IC-3).

Exercise 1.11.6

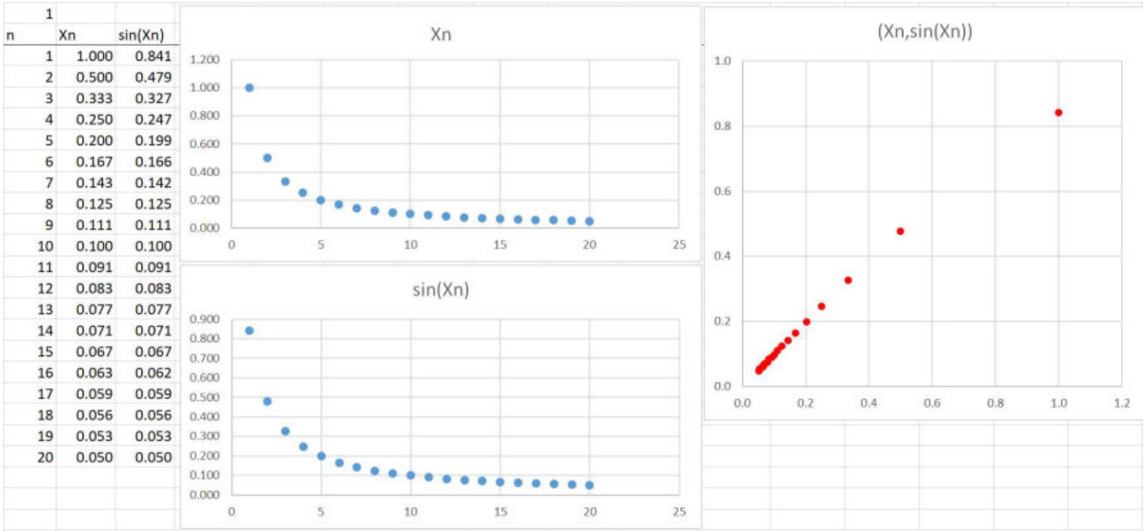
Prove the second half of the theorem.

There are more subtle observations about the graphs of these two functions.

First, the graph of $y = \sin x$ almost merges with the line $y = x$ around 0:



In fact, plotting the points $(1/n, \sin 1/n)$ reveals a straight line with slope 1:



	1	
n	Xn	sin(Xn)
1	1.000	0.841
2	0.500	0.479
3	0.333	0.327
4	0.250	0.247
5	0.200	0.199
6	0.167	0.166
7	0.143	0.142
8	0.125	0.125
9	0.111	0.111
10	0.100	0.100
11	0.091	0.091
12	0.083	0.083
13	0.077	0.077
14	0.071	0.071
15	0.067	0.067
16	0.063	0.062
17	0.059	0.059
18	0.056	0.056
19	0.053	0.053
20	0.050	0.050

This is how we state this fact algebraically:

Theorem 1.11.7: Famous Limit for Sine

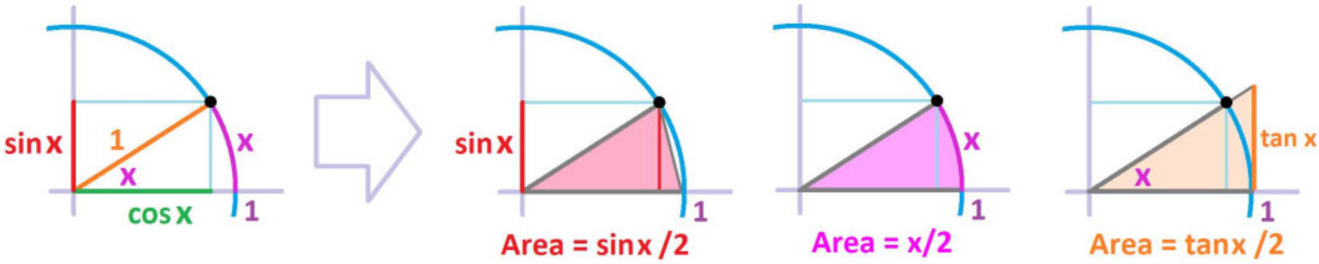
We have:

$$\boxed{\lim_{n\to\infty} \frac{\sin x_n}{x_n} = 1}$$

for any sequence $x_n \to 0$.

Proof.

The conclusion follows from certain trigonometry facts (seen in Volume 1, Chapter 1PC-4):



We can see three triangles and the following relation between their areas:

$$\text{red} \; < \; \text{purple} \; < \; \text{orange}$$

We use the following two facts:

- The area of a triangle with base 1 and height h is $h/2$.
- The area of a sector of the circle of radius 1 with angle x is $x/2$.

Therefore, we have:

$$\frac{1}{2}\sin x < \frac{1}{2}x < \frac{1}{2}\tan x \,,$$

or

$$\sin x < x < \tan x \,.$$

We flip it and multiply by $\sin x$:

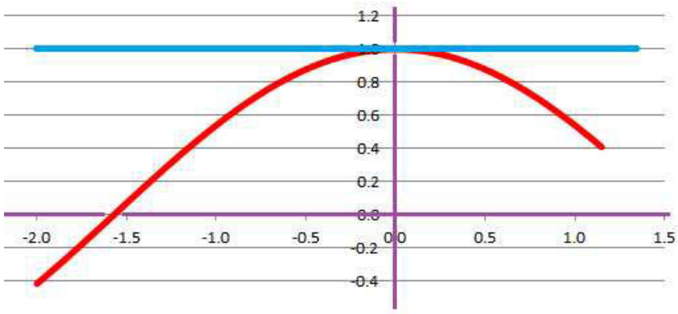
$$1 > \frac{\sin x}{x} > \cos x \,.$$

The sequence on the right $\cos x_n$ converges to 1 when $x_n \to 0$ by the last theorem. Therefore, so does the one in the middle by the *Squeeze Theorem*.

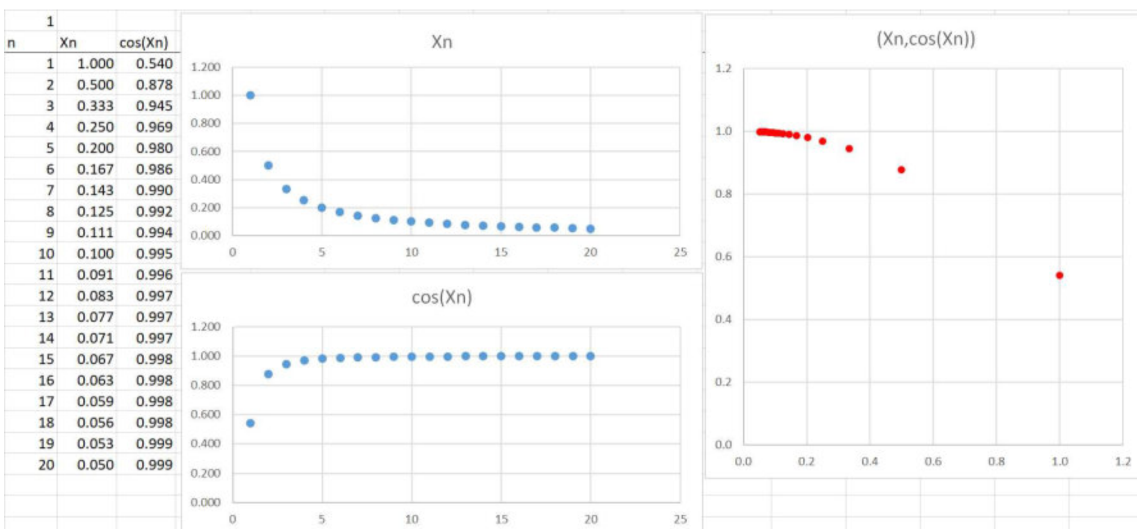
Warning!

We will still need to confirm that our application of the idea of "area" to a curved region is valid (Volume 3, Chapter 3IC-1).

Next, the graph of $y = \cos x$ almost merges with the line $y = 1$ when close to the y-axis:



Indeed, plotting the points $(1/n, \cos 1/n)$ shows that the slope converges to 0:



n	Xn	cos(Xn)
1	1.000	0.540
2	0.500	0.878
3	0.333	0.945
4	0.250	0.969
5	0.200	0.980
6	0.167	0.986
7	0.143	0.990
8	0.125	0.992
9	0.111	0.994
10	0.100	0.995
11	0.091	0.996
12	0.083	0.997
13	0.077	0.997
14	0.071	0.997
15	0.067	0.998
16	0.063	0.998
17	0.059	0.998
18	0.056	0.998
19	0.053	0.999
20	0.050	0.999

Let's compare the two algebraically:

Corollary 1.11.8: Famous Limit for Cosine

We have:

$$\boxed{\lim_{n\to\infty} \frac{\cos x_n - 1}{x_n} = 0}$$

for any sequence $x_n \to 0$.

Exercise 1.11.9

Prove the theorem.

Exercise 1.11.10

What does the graph of the tangent look like close to the y-axis?

An analogous result is true for the tangent:

Corollary 1.11.11: Famous Limit for Tangent

We have:

$$\boxed{\lim_{n\to\infty}\frac{\tan x_n}{x_n}=1}$$

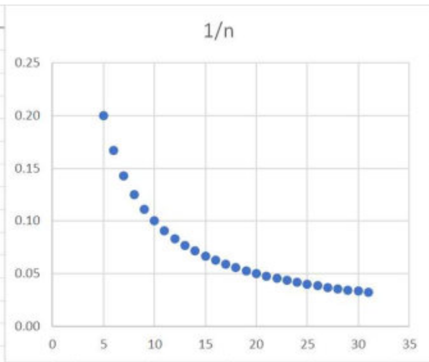
for any sequence $x_n \to 0$.

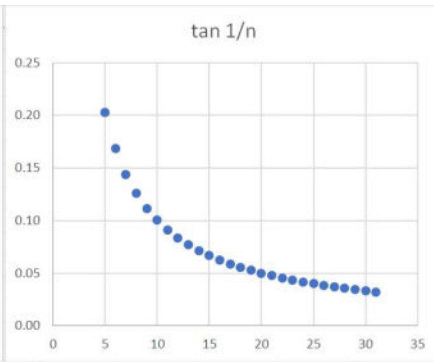
Proof.

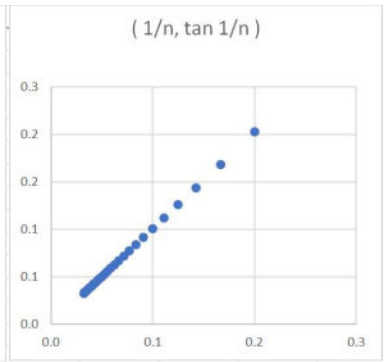
It follows from the above theorem, the fact that $\cos x_n \to 1$ for any sequence $x_n \to 0$, and the *Quotient Rule*.

This is a confirmation:

n	Xn=1/n	tan Xn
5	0.20	0.20
6	0.17	0.17
7	0.14	0.14
8	0.13	0.13
9	0.11	0.11
10	0.10	0.10
11	0.09	0.09
12	0.08	0.08
13	0.08	0.08
14	0.07	0.07
15	0.07	0.07
16	0.06	0.06
17	0.06	0.06
18	0.06	0.06
19	0.05	0.05







Chapter 2: Limits and continuity

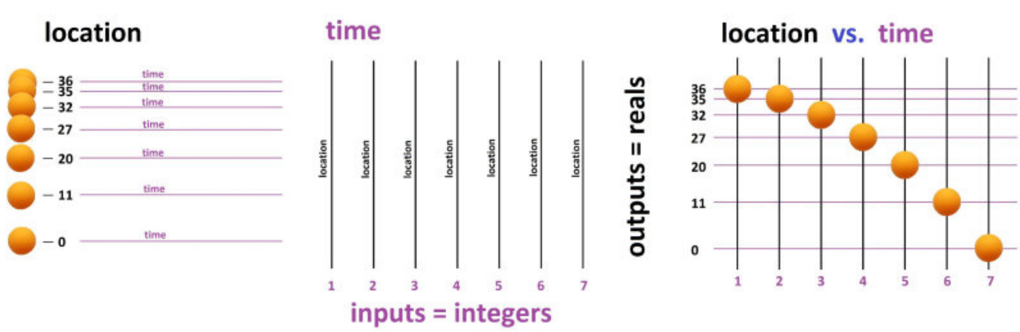
Contents

2.1. Functions

In Chapter 1, we presented a very brief introduction to calculus. We chose to enter through the study of motion.

As we commonly think of motion as a continuous progress through physical *space*, we think of space as infinitely divisible (i.e., the set of real numbers). If we also think of motion as an incremental progress through *time*, we think of time as discrete (i.e., the set of integers). The latter is the input and the former is the output of these functions that we call *sequences*!

This is how we visualize sequences:

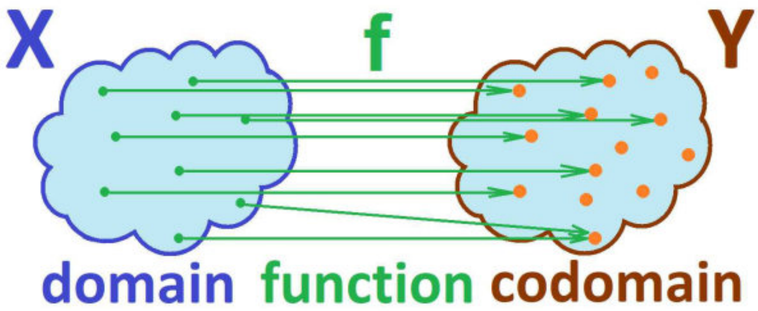


But how do we treat motion when the time varies continuously? This requires us to think of time as the real numbers, too. We need a different class of *functions*: real input and real output!

Warning!

In the long run, we will be deciding on multiple occasions what kinds of inputs and outputs we need.

Recall our general view on functions:



As we can see, there must be a y for each x, but not necessarily vice versa. The whole calculus is built on the following concept as a foundation:

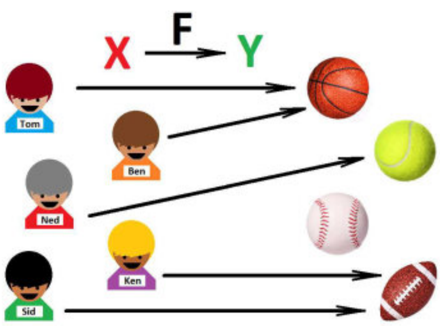
Definition 2.1.1: function

A *function* is a rule or procedure f that assigns to any element x in a set X, called the *input set* or the *domain* of f, exactly one element y, which is then denoted by

$$y = f(x)$$

in another set Y. The latter set is called the *output set* or the *codomain* of f. The inputs are collectively called the *independent variable*; the outputs are collectively called the *dependent variable*. We also say that *the value of x under f* is y.

The term "variable" is explained by the fact that if we can *vary* the inputs x within set X, the outputs y will *vary* within set Y. The idea has nothing to do with numbers. Here is an example of a function:



The arrows indicate what sport each boy prefers, and as we go down the list of boys on the left, the arrow automatically trace the balls on the right.

Below is the common notation:

Function from set to set

$$f : X \to Y$$

or

$$X \xrightarrow{\quad f \quad} Y$$

It reads "function f from X to Y".

In this volume we will use primarily the following two types of functions:

Definition 2.1.2: numerical functions and sequences

- A *numerical function* is a function from a subset of the real numbers, $X \subset \mathbf{R}$, to the real numbers:

$$f : X \to \mathbf{R}\,.$$

- A *sequence* is a numerical function with the domain X consisting of consecutive integers.

Let's compare:

- A sequence $a_n : X \to \mathbf{R}$.
  - The input variable is n, an integer ($X \subset \mathbf{Z}$).
  - The output variable is $y = a_n$, a real number.
- A numerical function $f : X \to \mathbf{R}$.
  - The input variable is x, a real number ($X \subset \mathbf{R}$).
  - The output variable is $y = f(x)$, another real number.

Warning!

The latter class includes the former.

We compare the notations too, side by side:

Function vs. sequence

name of the function

$$\begin{array}{ccc} \downarrow & & \downarrow \\ f(\ \ x\ \ ) & \text{vs.} & a\ \ {}_{n} \\ \ \ \uparrow & & \ \ \uparrow \end{array}$$

name of the input variable

value of the input variable

$$\begin{array}{ccc} \downarrow & & \downarrow \\ f(\ \ 3\ \ ) \ = 5 & \text{vs.} & a\ \ {}_{3}\ \ = 5 \\ \qquad\uparrow & & \qquad\uparrow \end{array}$$

value of the output variable

Graphs provide a way to have a bird's eye view of the function.

Recall that the *graph* of a function is the set of points in the xy-plane that satisfy t$y = f(x)$. We can express this idea with the the set-building notation (seen in Volume 1, Chapter 1PC-2):

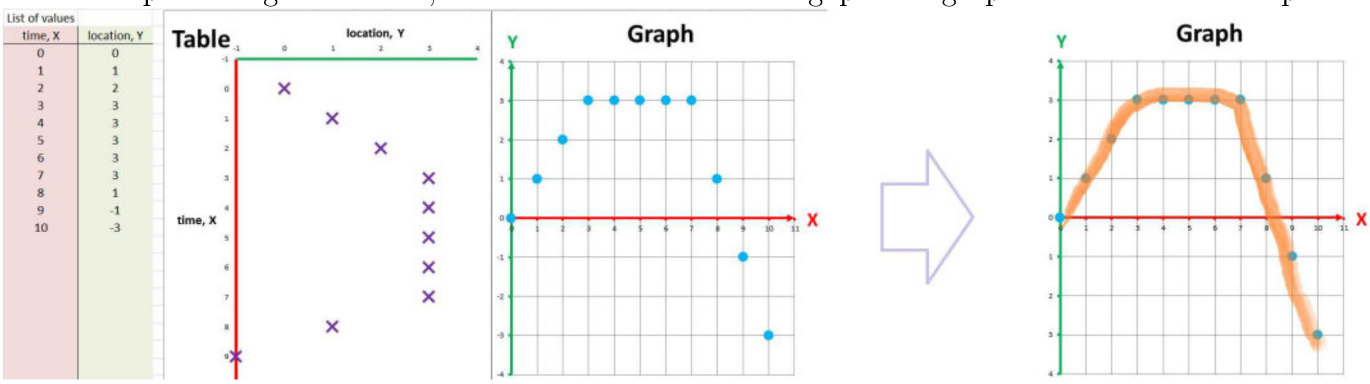
$$\boxed{\text{graph of } f \;=\; \{(x, y) :\; y = f(x)\}}$$

In other words, the graph is the set of all possible points on the plane in the form $(x, f(x))$.

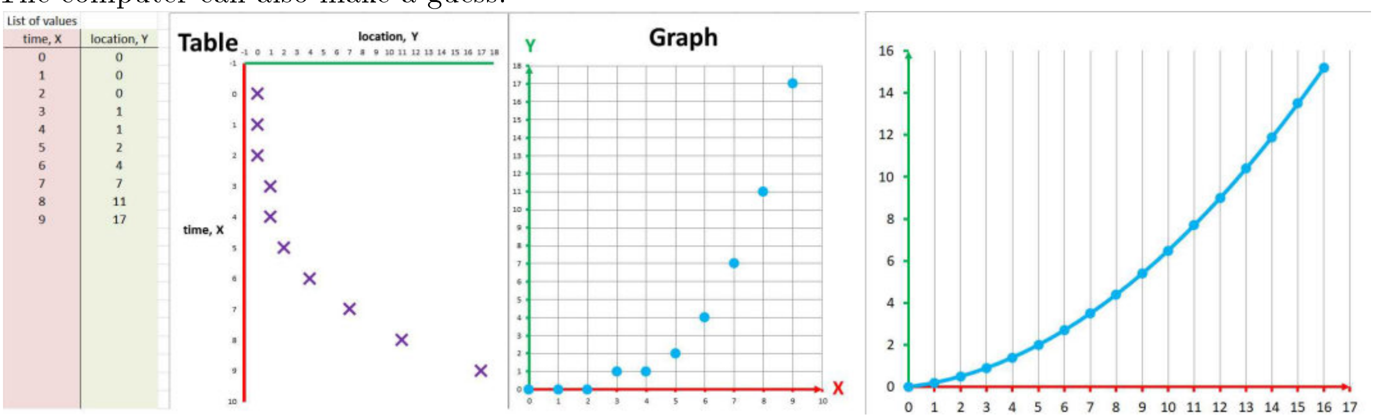
We insert more inputs as necessary. When there are enough of them, they start to form a curve!

Example 2.1.3: graphs as curves

Instead of producing more data, it is common to fill in the gaps in a graph with a stroke of a pen:

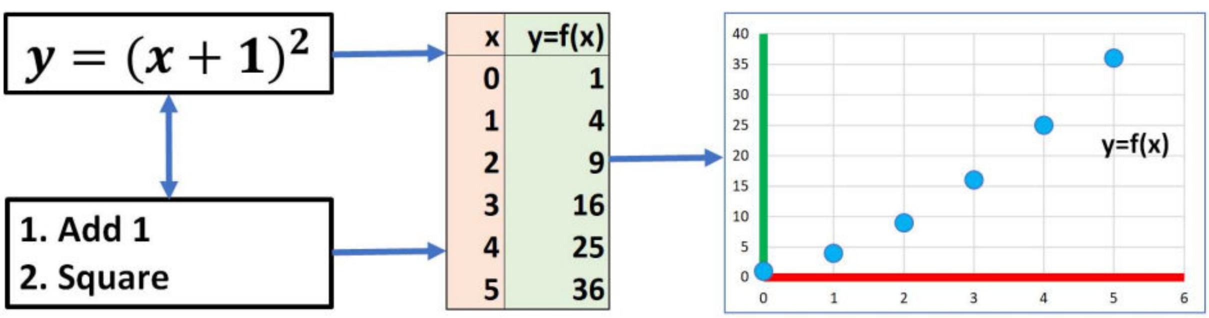


The computer can also make a guess:



Example 2.1.4: representations of functions

There are several ways to represent and visualize a function. They are interconnected:



In summary, we should move freely among these treatments of functions:

formula ↘
↕ list ⟷ graph
program ↗

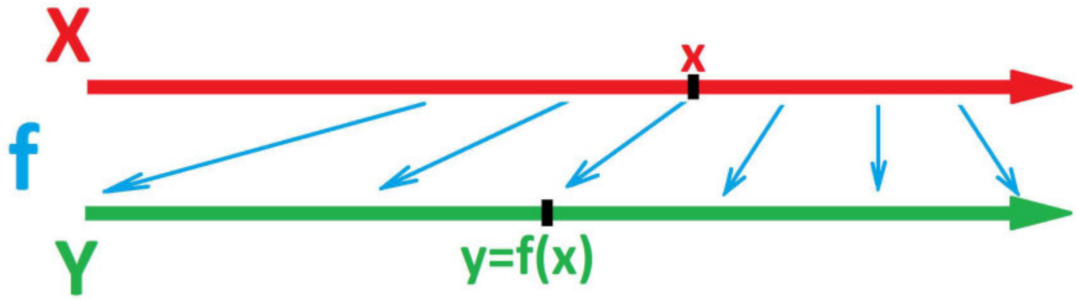
When a numerical function $y = f(x)$ is given to us without any prior background, we can still supply it with a *tangible representation*:

- We think of the function as if it represents *motion* (x time, y location).
- We plot the *graph* of the function on a piece of paper.
- We think of the function as a *transformation* of the real line.

Let's review the last item (seen in Volume 1, Chapter 1PC-3):

▶ Numerical functions are transformations of the real number line... and vice versa.

We represent a numerical function as a correspondence between the x- and the y-axis:

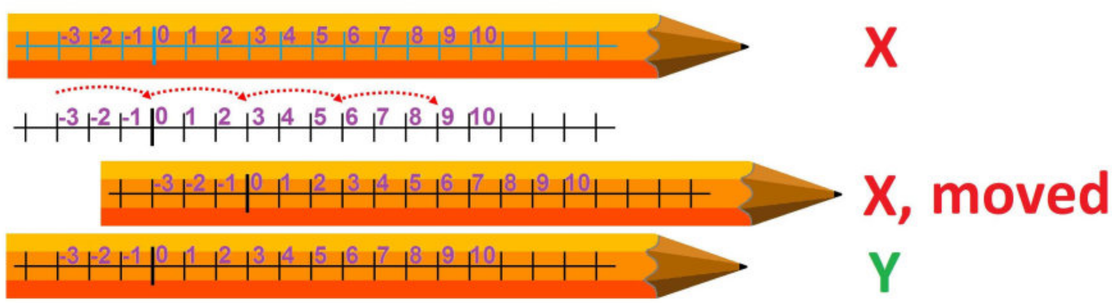


In contrast to the xy-plane that contains the graph, they are parallel rather than perpendicular to each other!

For example, the function given by

$$y = x + s \quad \text{and} \quad x \ \mapsto \ x + s$$

shifts the x-axis in the positive direction when $s > 0$, and in the negative direction when $s < 0$:

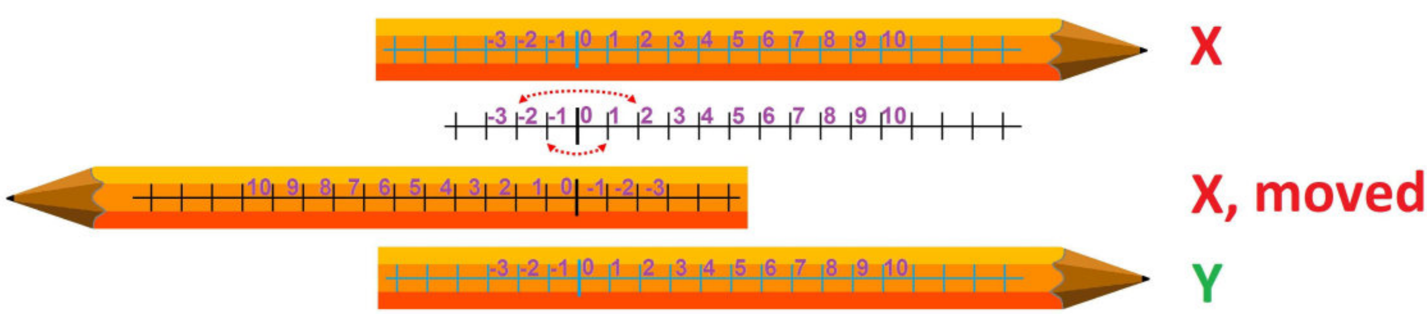


It's a rigid motion.

Next, the function given by

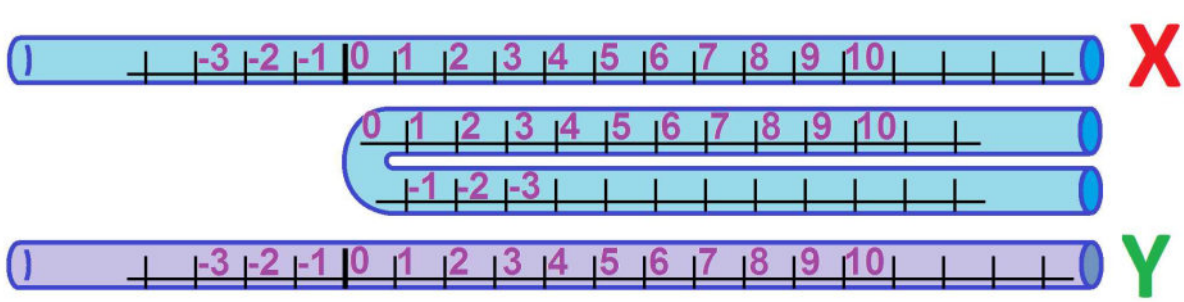
$$y = -x \quad \text{and} \quad x \ \mapsto \ -x$$

interchanges two points on the x-axis:



This is also a rigid motion. What these transformations have in common is that each takes the x-axis by two spots and then bring those two to the assigned locations on the y-axis.

Consider now the *folding* of a piece of wire:



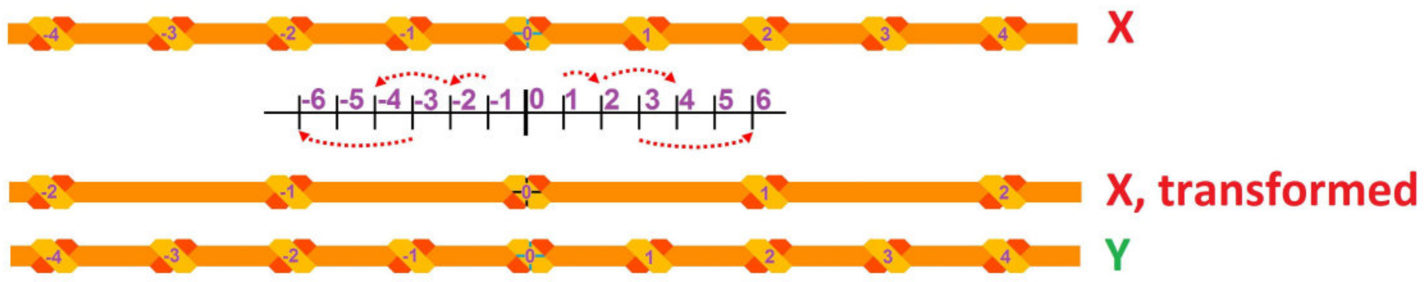
This transformation is given by the absolute value function:

$$y = |x| \quad \text{and} \quad x \mapsto |x|\,.$$

Another type of transformation is the *stretch* of the x axis as if it's a rubber string, given by

$$y = x \cdot k \quad \text{and} \quad x \to x \cdot k\,.$$

We grab it by the ends and pull them apart:

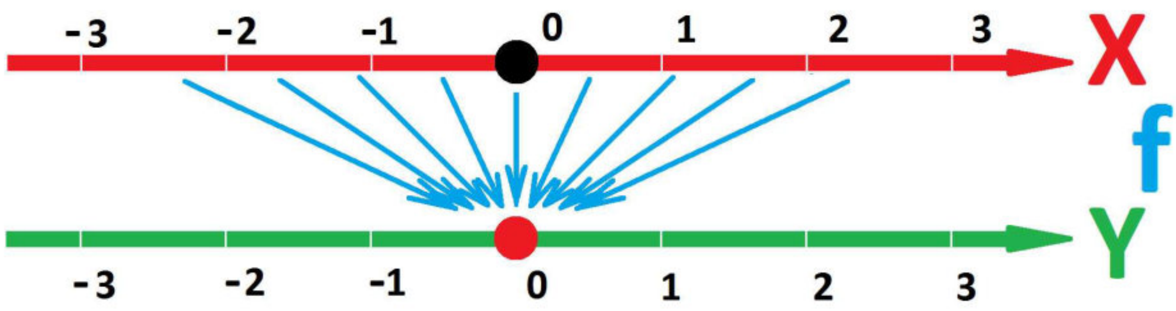


This is indeed a uniform stretch because the distance between *any* two points doubles. We understand "stretched by a factor k" as "shrunk by a factor $1/k$".

A constant function, given by

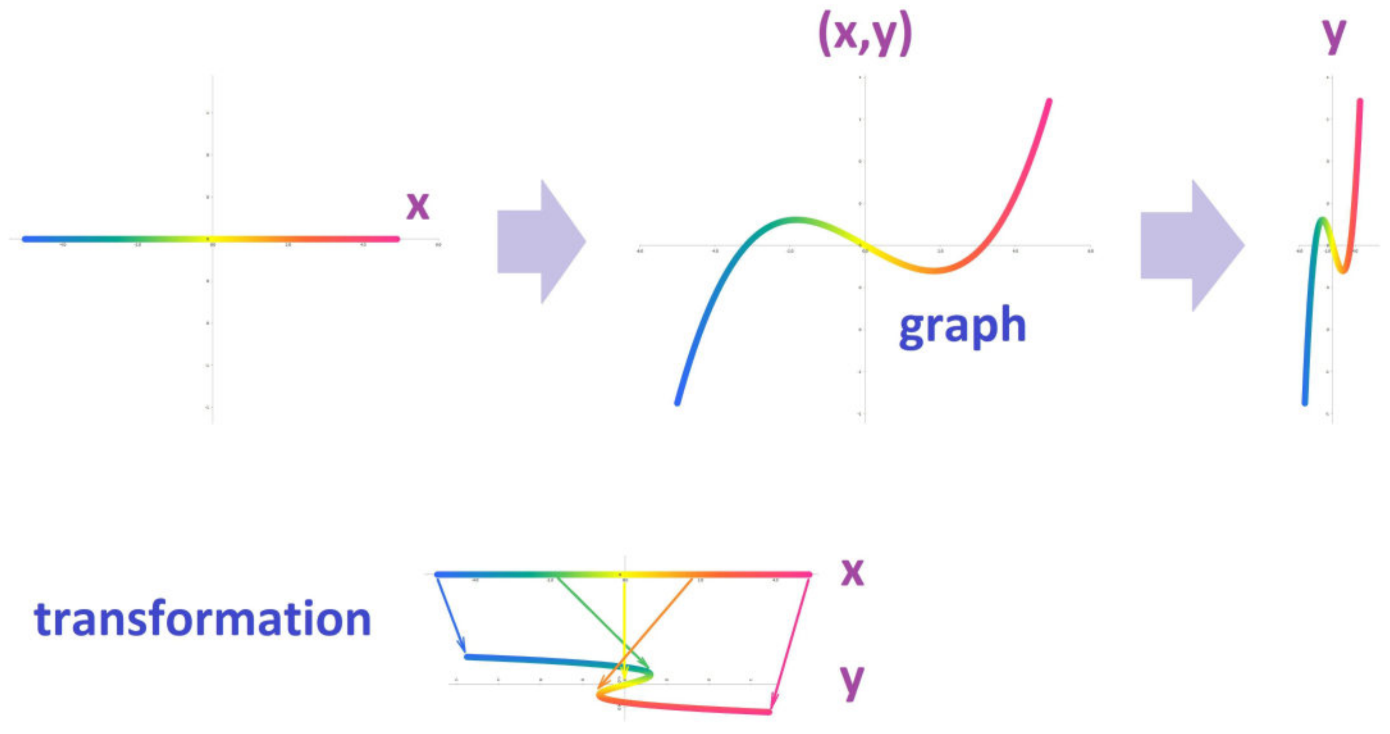
$$f(x) = c \quad \text{and} \quad x \mapsto c\,,$$

is seen as a *collapse*.



It shrinks to a single point.

The transformation of the domain into the codomain performed by the function can also be seen in its graph:



First, we take the x-axis as if it is a rope and lift it vertically to the graph of the function and then we push it horizontally to the y-axis.

Exercise 2.1.5

Classify these functions:

function	odd	even	onto	one-to-one
$f(x) = 2x - 1$				
$g(x) = -x + 2$				
$h(x) = 3$				

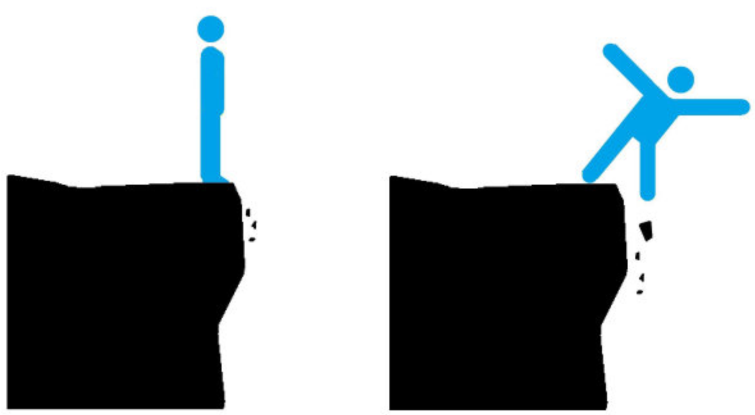
Exercise 2.1.6

Explain the difference between these two functions:

$$\sqrt{\frac{x-1}{x+1}} \text{ and } \frac{\sqrt{x-1}}{\sqrt{x+1}}\,.$$

2.2. Continuity and discontinuity

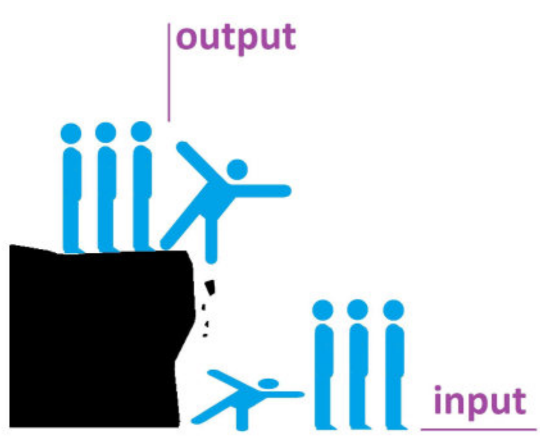
If one stands at the edge of a cliff, one small step may lead to catastrophic consequences:



Standing off the edge will not. One can capture this phenomenon with the graph of this function:

$$f(x) = \begin{cases} 1 & \text{if } x \le 0\,, \\ 0 & \text{if } x > 0\,. \end{cases}$$

Here x is the person's horizontal position and $f(x)$ is the vertical position:

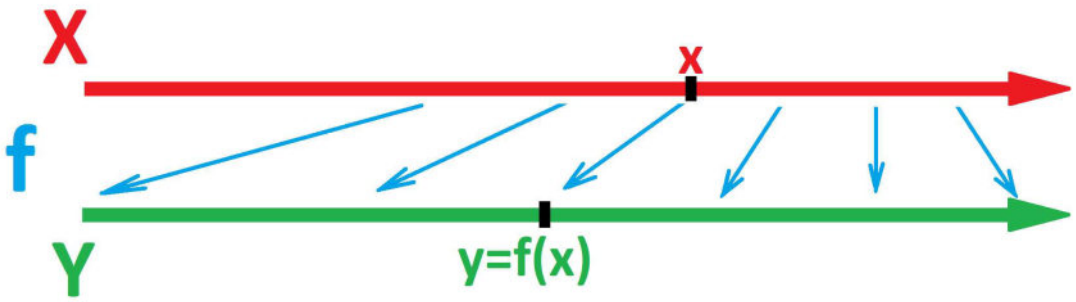


No matter how small a step (in the positive direction) is – $1, 1/2, 1/3, ..., 1/n, ...$ – it is fatal.

It is called *discontinuity*.

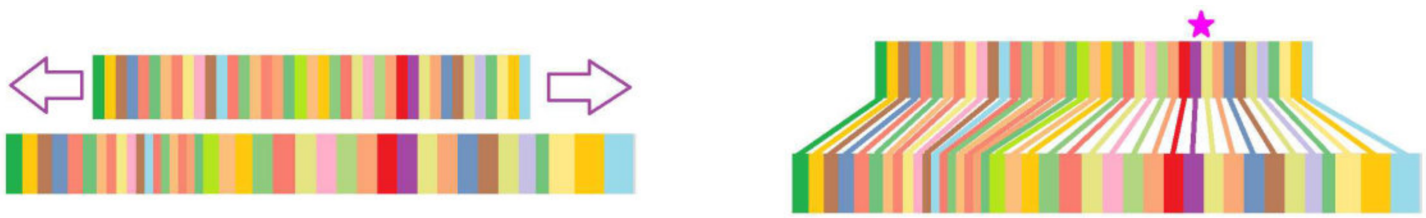
In this volume, we will study functions on a *small* scale. That is why the functions that may produce a dramatically different output in response to a small change in the input will not initially be a subject of our study. In Volume 3, we will turn to the study of functions on a *large* scale and such functions won't be as big a concern.

The idea how a tiny change can make a difference becomes especially vivid when we look at numerical functions as *transformations*:



The arrows not only tell us what happens to each number but they also suggest what happens to the *whole* x-axis!

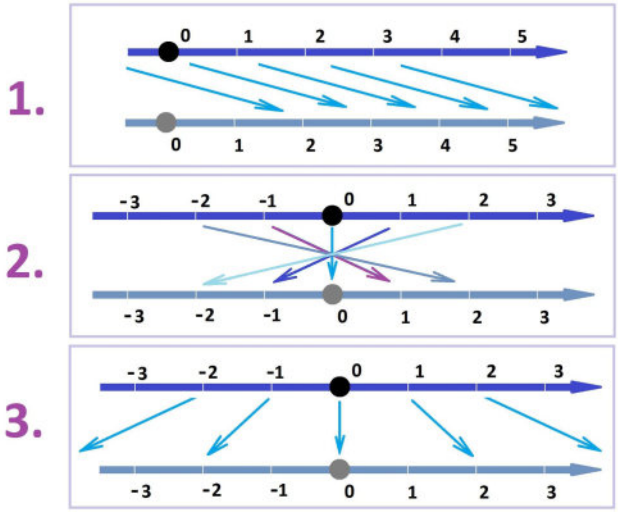
Here we see a variety of shifting, stretching, and shrinking at different locations:



There is no tearing though!

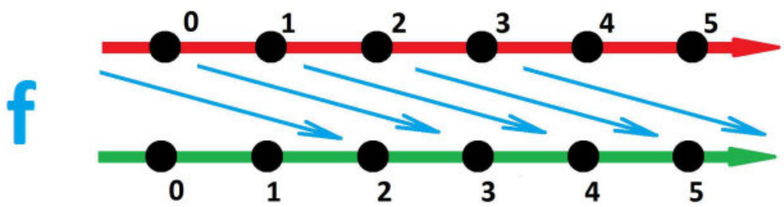
Example 2.2.1: intact x-axis

The basic transformations in the last section are: shifting, stretching, flipping. The points are moved in unison:

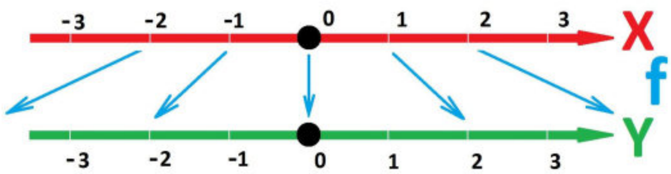


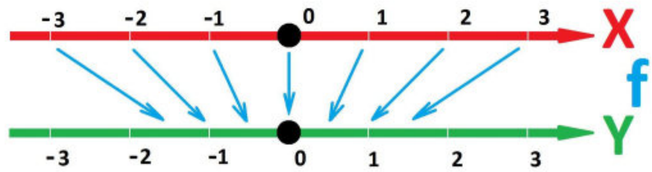
We can make this idea more precise.

In a shift, if x changes by p, then so does y:



This is a stretch by 2 and a shrink by 2:

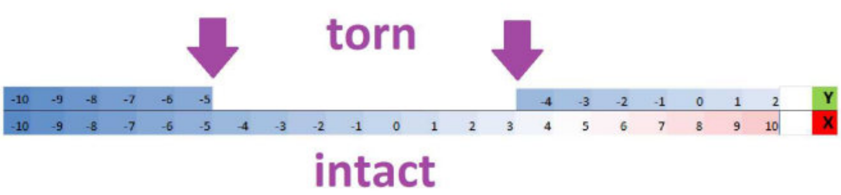


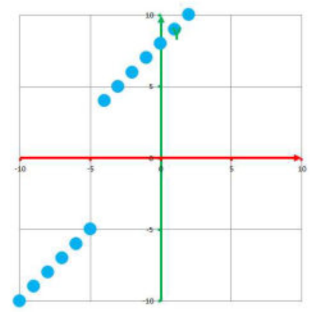


If x changes by p, then y changes by $2p$. And so on.

Example 2.2.2: tearing of x-axis

The x-axis might be torn by the transformation:





We can see how the color of the output changes abruptly at one location! Meanwhile, there is a gap in the graph – discontinuity.

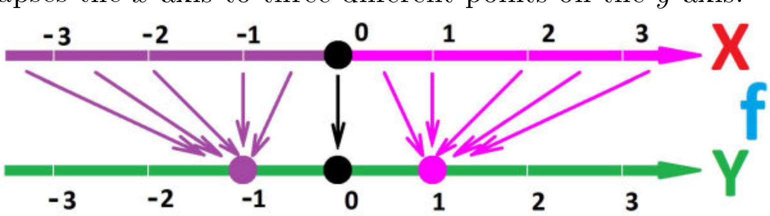
Let's look at the formula:

$$f(x) = \begin{cases} x+8 & \text{when } x > -5\,, \\ x & \text{when } x \le -5\,. \end{cases}$$

The function has a special point: No matter how small a deviation of x from $a = -5$ is, it will produce a large deviation of the output from $b = f(a)$ to $y = f(x)$.

Example 2.2.3: sign function

The sign function collapses the x-axis to three different points on the y-axis:



This may be our main observation about this function: Even the smallest step away from 0 will change the value of the function from $f(0) = 0$ to $y = f(x) = \pm 1$.

Exercise 2.2.4

Discuss the continuity of the integer value function.

We have considered some "poorly behaved", discontinuous functions. Such a function would have at least one special point:

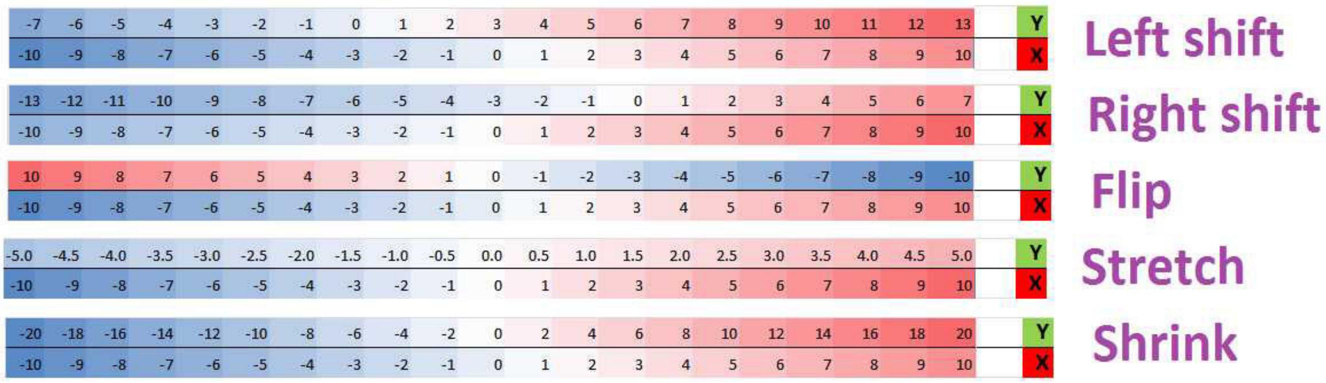
- ▶ No matter how small a deviation of the input x from a is, it will produce a large deviation of the output $y = f(x)$ from $b = f(a)$.

So, what is the opposite? This is what we mean by a *continuous dependence* of y on x under function f:

▶ A small deviation of the input x from a will produce a small deviation of the output $y = f(x)$ from $b = f(a)$.

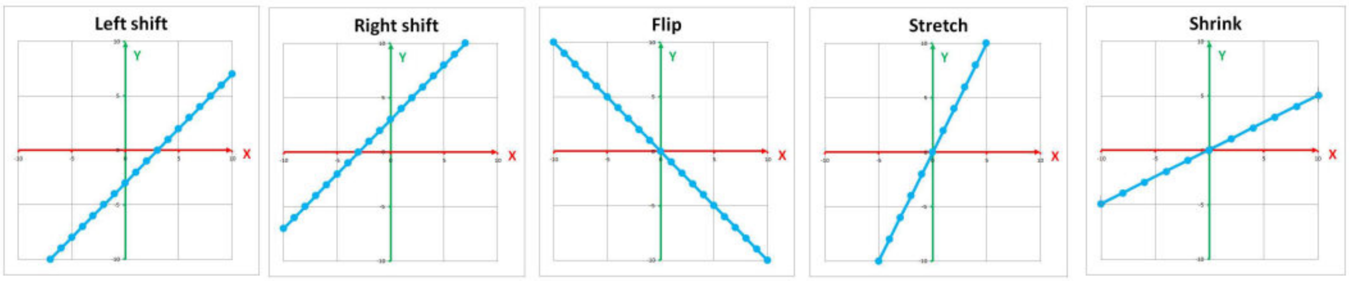
Example 2.2.5: basic transformations

The idea is visualized below:



The color of the output never changes abruptly!

We also plot the graphs of these functions below:



There are no gaps!

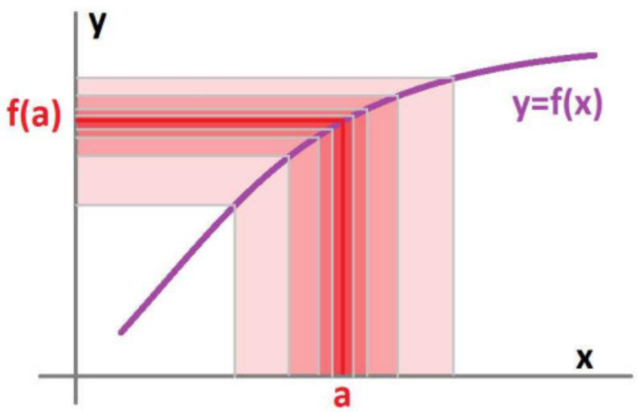
Continuous dependencies are ubiquitous in nature:

- The location of a moving object continuously depends on time.
- The pressure in a closed container continuously depends on the temperature.
- The air resistance continuously depends on the velocity of the moving object, etc.

It is our goal to develop this idea in full and to make sure that our mathematical tools match the perceived reality.

We take a better look at the *graphs*.

We make visible the idea that a small deviation of x from a will produce a small deviation of of $y = f(x)$ from $b = f(a)$:

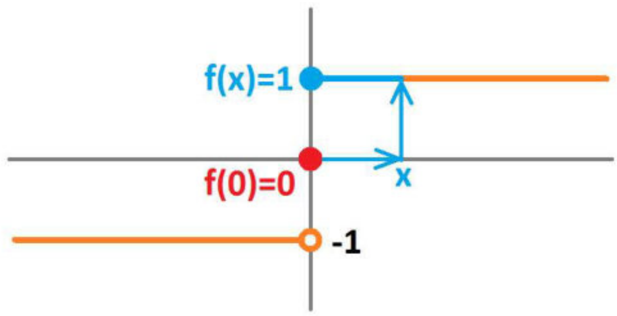


We can see several (smaller and smaller) deviations of x and the corresponding deviations of y (also smaller and smaller).

It is revealed that our mathematical tool will be *limits* of sequences!

Example 2.2.6: sign function

The sign function $f(x) = \text{sign}(x)$ has a visible *gap* at $x = 0$:



This example shows how the idea of continuous dependence of y on x fails: Starting with $x = 0$, even the tiniest deviation of x, say, $x = .0001$, produces a jump in y from $\text{sign}(0) = 0$ to $\text{sign}(.0001) = 1$. We interpret this observation via limits. Let's consider the *compositions* of f with a few sequences that converge to 0:

$$\lim_{n\to\infty} x_n = 0\,.$$

First, try $x_n = 1/n$. We have:

$$\lim_{n\to\infty} \text{sign}\left(\frac{1}{n}\right) = \lim_{n\to\infty} 1 = 1\,.$$

However, what about $x_n = -1/n$? We have:

$$\lim_{n\to\infty} \text{sign}\left(-\frac{1}{n}\right) = \lim_{n\to\infty} (-1) = -1\,.$$

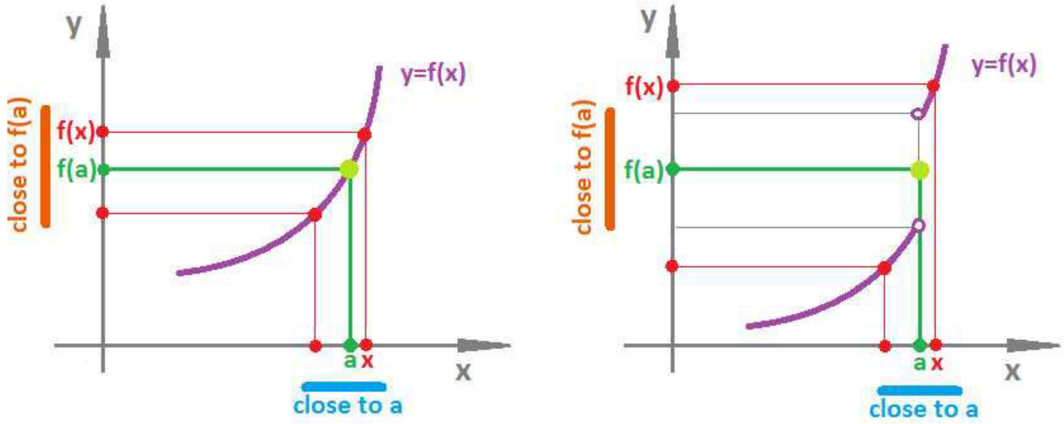
The limits don't match! Furthermore, the following limit doesn't even exist as we know from Chapter 1:

$$\lim_{n\to\infty} \text{sign}\left(\frac{(-1)^n}{n}\right) = \lim_{n\to\infty} (-1)^n, \quad \text{no limit.}$$

All of this points at discontinuity!

Example 2.2.7: jump discontinuity

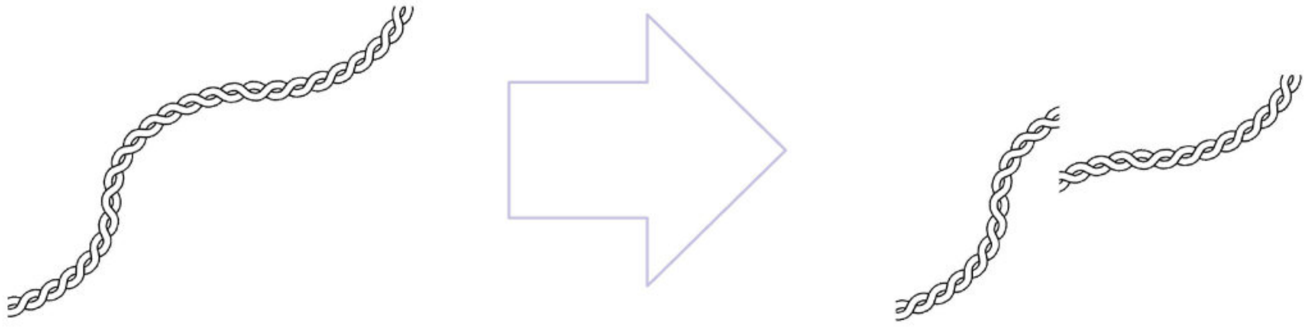
A more general view of the "jump discontinuity" is shown below:



On the right, these x's that are "close" to a don't all produce y's "close" to $f(a)$.

Example 2.2.8: rope

One of the informal ways to speak about continuity is to say that the graph of such a function is "made of a single piece". We can even think of the graph as a rope:



Even though it looks intact, there may be an invisible cut that allow us to pull it apart.

Example 2.2.9: cuts in graphs

Let's investigate the following two functions around $a = 1$:

$$f(x) = \frac{x^2 - 1}{x - 1} \quad \text{and} \quad g(x) = x + 1\,.$$

Suppose we have a sequence $x_n \to 1$ as $n \to \infty$ (such as $1 + 1/n$). These are the x's. What happens to the corresponding y's?

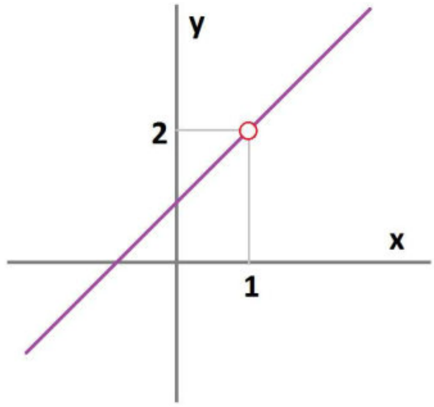
The limit of the former is found by this computation:

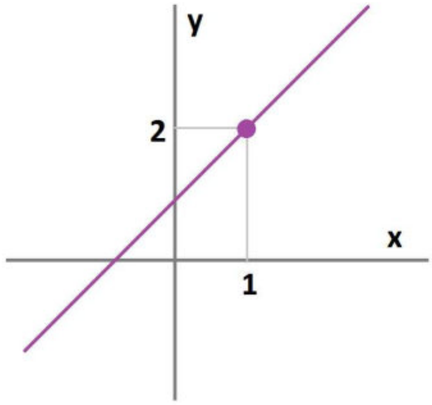
$$\begin{aligned}
\lim_{n\to\infty} f(x_n) &= \lim_{x\to 1} \frac{x_n^2 - 1}{x_n - 1} && \to \frac{0}{0}\,? \quad \boxed{\textbf{DEAD END}} \\
&= \lim_{n\to\infty} \frac{(x_n - 1)(x_n + 1)}{x_n - 1} \\
&= \lim_{n\to\infty} (x_n + 1) \\
&= \lim_{n\to\infty} x_n + 1 \\
&= 2\,.
\end{aligned}$$

The limit of the latter is found too:

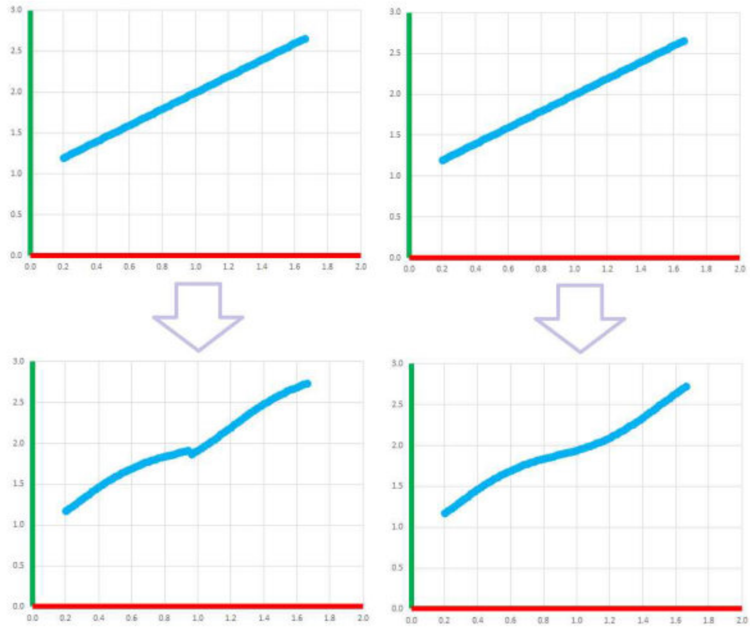
$$\lim_{n\to\infty} g(x_n) = \lim_{n\to\infty} (x_n + 1) = \lim_{n\to\infty} x_n + 1 = 1 + 1 = 2\,.$$

The two functions are almost the same, and the difference is seen in their graphs below. To emphasize the difference, we use a little circle to indicate the missing point:





There is only one point missing from the former graph. However, if we think of the graph of a function as a rope, we realize that the former graph consists of two separate pieces! Even though the cut is invisibly thin, we can pull the pieces apart. The latter graph is a single piece; it is indeed "continuous". A light breeze would blow apart the former but only move the latter:

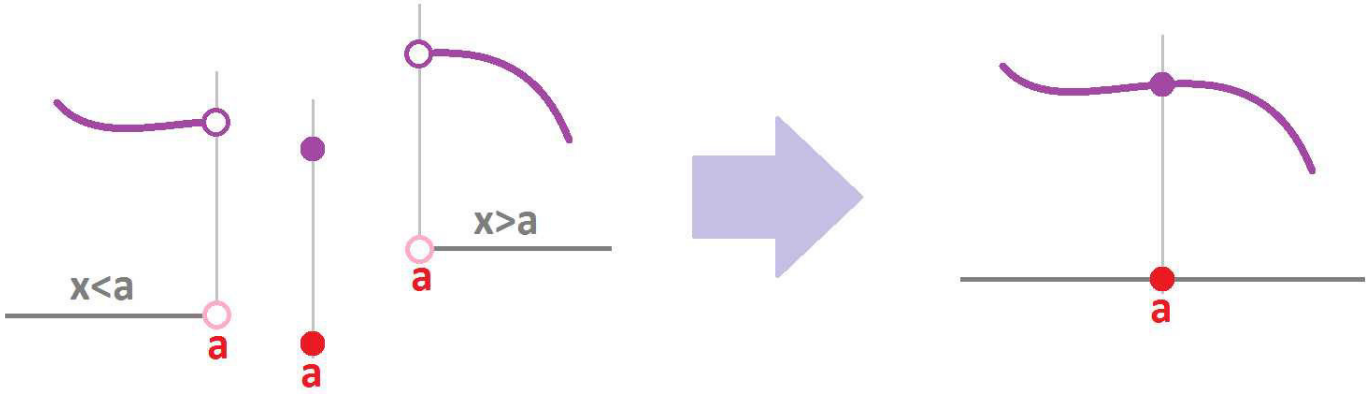


The discontinuous graph in the last example is easy to "repair" (glue, solder, weld, etc.) by adding a single point. The jump discontinuity is a more extreme, case of discontinuity; it cannot be fixed.

The examples teach us a lesson. Given a function f and a point a, where f is defined, the graph of f consists of three parts:

1. the part of the graph of f with $x < a$,
2. the part of the graph of f with $x = a$ (one point), and
3. the part of the graph of f with $x > a$.

For this function to be continuous, these three parts (the two pieces of the rope and a drop of glue) have to fit together:



How? We require that the following two types of limits exist in the first place:

$$\lim_{n\to\infty} f(x_n) = f(a) \quad \text{for each sequence} \quad x_n \to a \quad \text{with} \quad \begin{array}{c} x_n < a \\ x_n > a \end{array} .$$

And then they have to be equal to the value of the function at a.

We can guess that the sequences that alternate between the sides will also have this property. So, for simplicity, we include all sequences that converge to a in any manner:

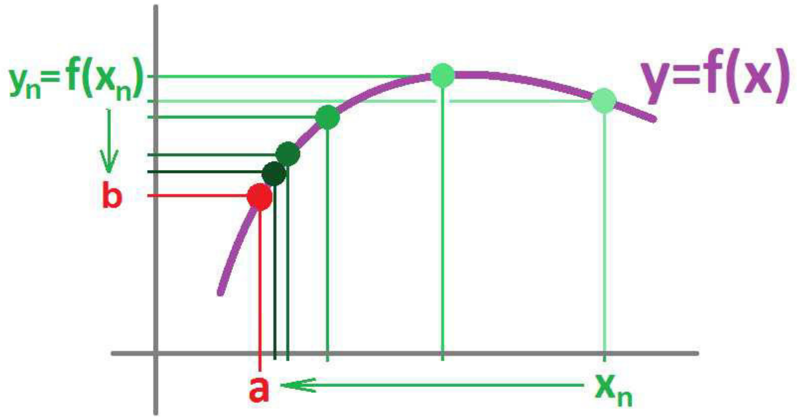
Definition 2.2.10: continuity

A function f is called *continuous at a point* $x = a$ when f is defined at $x = a$ and the limit below exists and equal to the value of the function at a:

$$\boxed{\lim_{n\to\infty} f(x_n) = f(a)}$$

for each sequence $x_n \to a$. If the limit doesn't exist, the function is called *discontinuous*.

The function is "sampled" by a sequence of inputs to see what happens to the outputs:



Warning!

Continuity is established one point at a time.

Example 2.2.11: logic

The definition includes the following phrase:

▶ `FOR ANY` sequence the condition is satisfied.

Therefore, the definition of discontinuity will include:

▶ `THERE IS` a sequence for which the condition is `NOT` satisfied.

In other words, a "test" sequence is chosen and subjected to the computation in the definition. If it fails, the function is discontinuous. If it passed, we will have to try another, and maybe all of them.

Where do discontinuous functions come from?

We can observe them in nature around us:

- We walk continuously – until we fall off a cliff.
- We continuously increase the pressure in a closed container – until it explodes.
- We continuously increase the temperature of a piece of ice – until it melts.

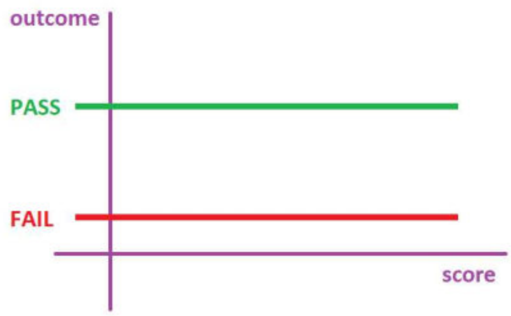
Examples of discontinuity routinely come from human affairs because the outcomes are often designed to be *discrete*: Yes or No, all or nothing, etc.

Example 2.2.12: grades

Let's consider possible outcomes in a university environment. We start with pass/fail. The outcome depends on your total (or average) score. A threshold is typically assigned: If the score is above it, you pass; otherwise you fail no matter how close you are. That's discontinuity!

Using letter grades relieve this somewhat, but there are still thresholds! They will remain even if we

choose to use $A-$, $B+$, etc.

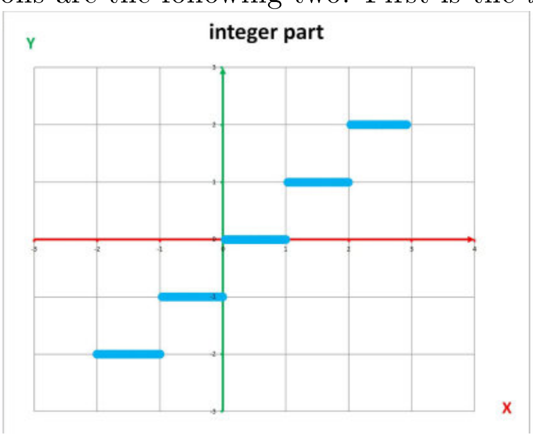


Generally, there are only a few possible outputs (the letter grades) but the input (the score) is an arbitrary real number. To try to plot the graph of how the outcomes depend on the score, we show above the lines that *must* contain the graph of the function. There cannot be a continuous function with such a graph.

Example 2.2.13: integer and fractional part

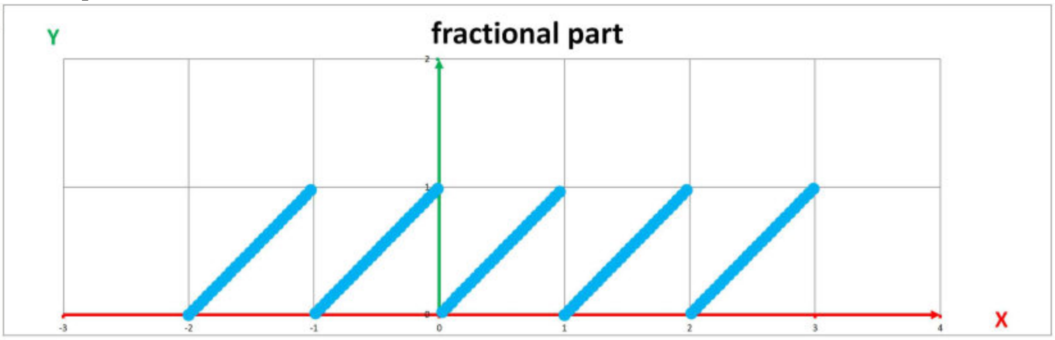
A pair of *useful* discontinuous functions are the following two. First is the *integer part*:

x	y=[x]
-2.00	-2
-1.96	-2
-1.93	-2
-1.89	-2
-1.85	-2
-1.82	-2
-1.78	-2
-1.74	-2
-1.70	-2
-1.67	-2
-1.63	-2
-1.59	-2
-1.56	-2
-1.52	-2
-1.48	-2
-1.45	-2
-1.41	-2
-1.37	-2
-1.33	-2
-1.30	-2
-1.26	-2



The second is the *fractional part*:

x	y=[x]	y={x}
-2.00	-2	0.00
-1.96	-2	0.04
-1.93	-2	0.08
-1.89	-2	0.11
-1.85	-2	0.15
-1.81	-2	0.19
-1.78	-2	0.23
-1.74	-2	0.26
-1.70	-2	0.30
-1.66	-2	0.34
-1.63	-2	0.38
-1.59	-2	0.41



Example 2.2.14: income tax

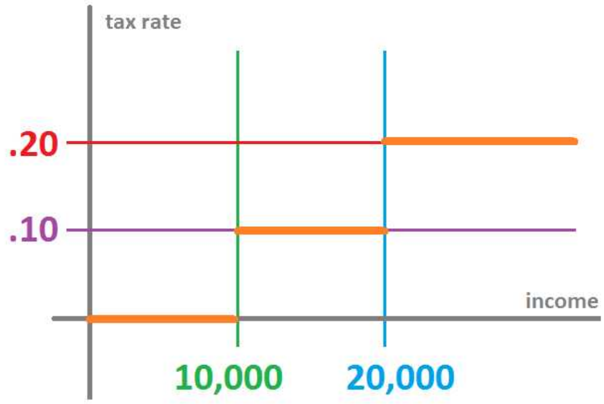
A special care has to be taken in order to ensure continuity. Let's consider this example that deals with taxes (Volume 1, Chapter 1PC-4). Hypothetically, suppose the tax code says about these three brackets of income:

- If your income is less than \$10000, there is no income tax.
- If your income is between \$10000 and \$20000, the tax rate is 10%.
- If your income is over \$20000, the tax rate is 20%.

We can express this algebraically. Suppose x is the income and $y = f(x)$ is the tax rate, then

$$f(x) = \begin{cases} 0 & \text{if } x \le 10000\,, \\ .10 & \text{if } 10000 < x \le 20000\,, \\ .20 & \text{if } 20000 < x\,. \end{cases}$$

The function is, of course, discontinuous (left):



We see why this is a problem once we try to apply this formula. Imagine that your income has risen from $\$10,000$ to $\$10,001$. Then your tax bill rises from $\$0$ to $\$1,000$!

How do we fix this? We need to assure that an increase in the income will cause only a smaller increase in the tax bill. This requirement implies the principle we have discussed: A "small" change in the input (the income) will cause only a "small" change in the output (the tax bill).

We notice that this requirement is satisfied as long as the income stays within the brackets. The issue arises only at the transition points and can be addressed in the way shown on the right. The interpretation of the result is possible if we understand the law correctly. The tax rate is "marginal"; i.e., it is the tax rate applied to the part of the income that lies within the bracket and these three numbers are meant to be added together to produce your tax bill liability.

This is the formula for the tax bill as a function of income:

$$g(x) = \begin{cases} 0 & \text{if } x \le 10000\,, \\ .10 \cdot (x - 10000) & \text{if } 10000 < x \le 20000\,, \\ .10 \cdot (x - 10000) + .20 \cdot (x - 20000) & \text{if } 20000 < x\,. \end{cases}$$

The function is continuous. Indeed, we have:

- If $x_n \to 10000$ and $x \le 10000$, then $g(x_n) = 0 \to 0\,.$
- If $x_n \to 10000$ and $x > 10000$, then $g(x_n) = .1(x_n - 10000) \to 0\,.$

Exercise 2.2.15

Prove that the formula above satisfies the requirements. Provide the missing parts of the proof of continuity.

Exercise 2.2.16

Provide a similar analysis for the possibility of a welfare payment of $\$1000$ to a person with an income below $\$5000$.

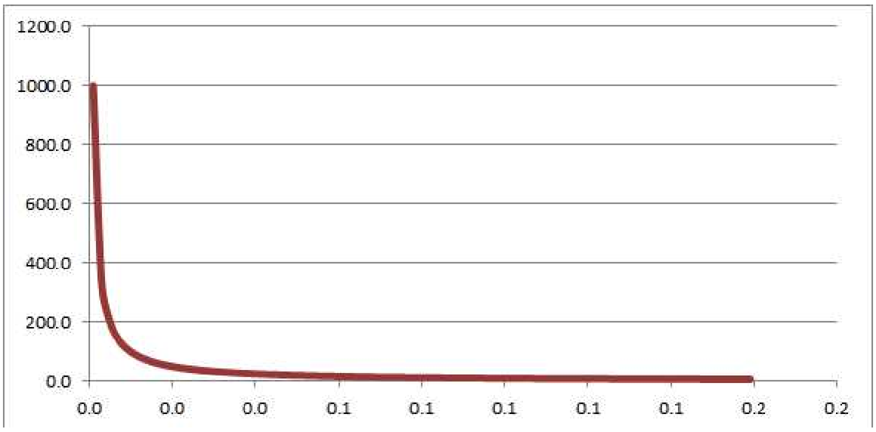
Even more extreme examples are below.

Example 2.2.17: reciprocal at 0

The reciprocal function

$$g(x) = \frac{1}{x}$$

has an *infinite gap* at $x = 0$:



Here, if the deviations of x from 0 differ in sign, the jump in y may be very large:

$$\text{from } \frac{1}{-.0001} = -10000 \text{ to } \frac{1}{.0001} = 10000\,.$$

A simple test sequence is $x_n = 1/n$. The composition with the function gives us the following:

$$y_n = g(x_n) = 1/x_n = 1/(1/n) = n \to +\infty\,.$$

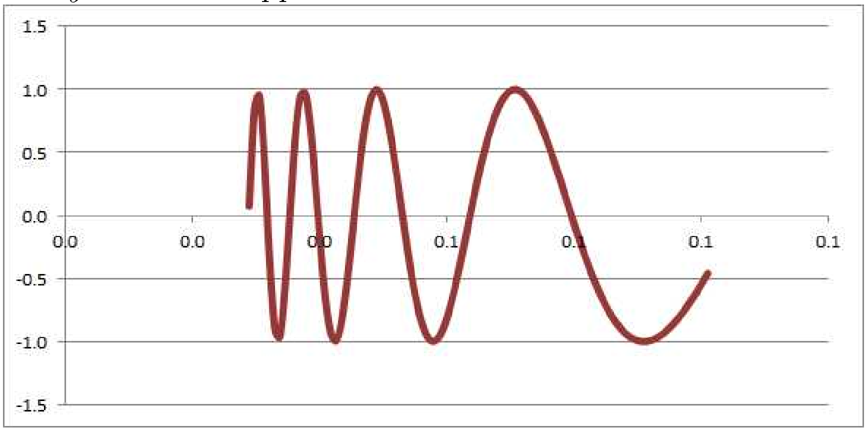
The discontinuity is proven.

Example 2.2.18: $\sin(1/x)$ **at** 0

The sine of the reciprocal

$$g(x) = \sin\left(\frac{1}{x}\right)$$

oscillates infinitely many times as x approaches 0:



Here, a deviation of x from 0 may unpredictably produce any number between -1 and 1. We can guess that the function cannot be continuous at 0!

To appreciate this conclusion from the geometric point of view, it is impossible to attach this graph to a single point on the y-axis no matter what that point may be. A simple test sequence is $x_n = 1/n$. The composition with the function gives us the following:

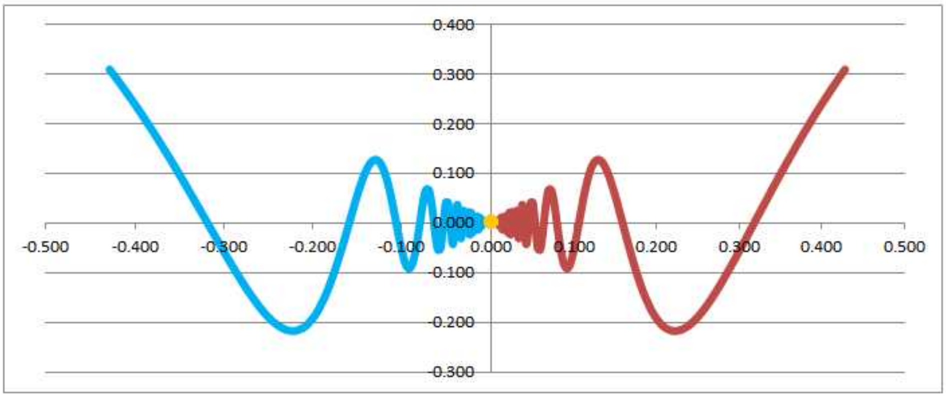
$$\lim_{n\to\infty} y_n = \lim_{n\to\infty} g(x_n) = \lim_{n\to\infty} \sin\left(\frac{1}{1/n}\right) = \lim_{n\to\infty} \sin n \quad \text{DNE.}$$

The discontinuity is proven.

In the meantime, the function

$$h(x) = x\sin(1/x)$$

also oscillates infinitely many times as we approach 0. Yet, because the magnitude of these oscillations diminishes, the limit of the composition with any test sequence exists (it's 0!):



Does this make the function continuous? No, there is still a point missing at 0. In other words, the two oscillating branches are detached. Let's glue them together. We simply define a new function:

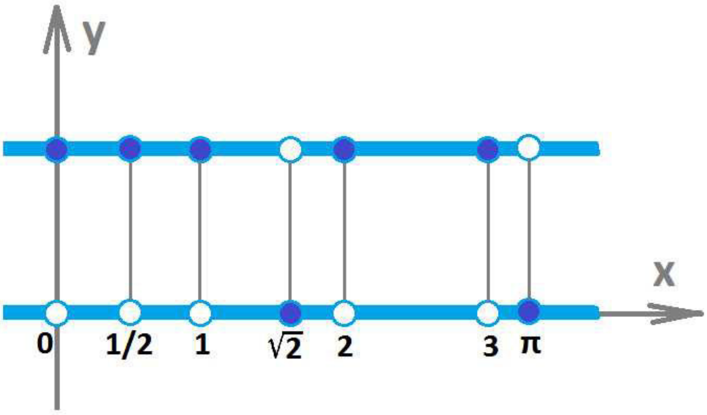
$$f(x) = \begin{cases} x\sin(1/x) & \text{if } x \neq 0\,, \\ 0 & \text{if } x = 0\,. \end{cases}$$

Example 2.2.19: nowhere continuous

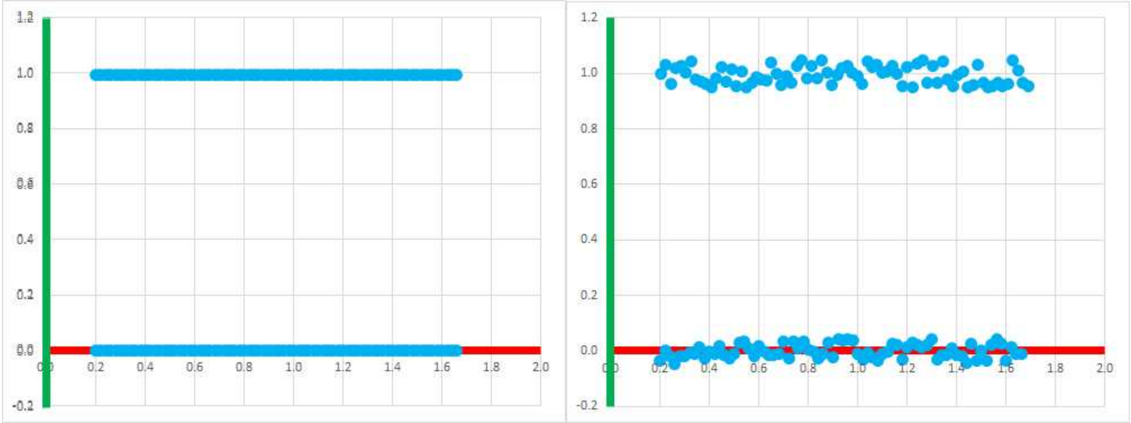
The *Dirichlet function* is nowhere continuous! It is defined by:

$$I_Q(x) = \begin{cases} 1 & \text{if } x \text{ is a rational number,} \\ 0 & \text{if } x \text{ is an irrational number.} \end{cases}$$

In an attempt to plot its graph, we can only draw two horizontal lines and then simply point out *some* of the missing points in either:

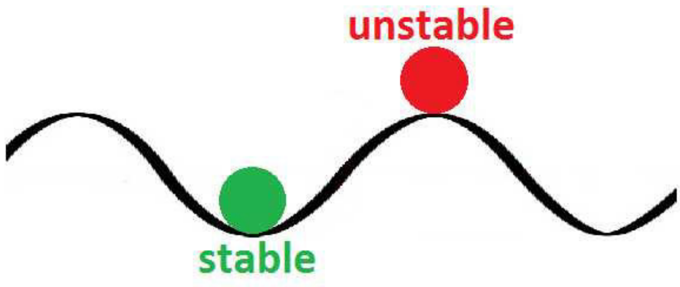


If we were to plot *all* points, we'd have what looks like two complete straight lines. However, a light breeze would blow them apart:



Exercise 2.2.20

Discuss continuity in the following context:

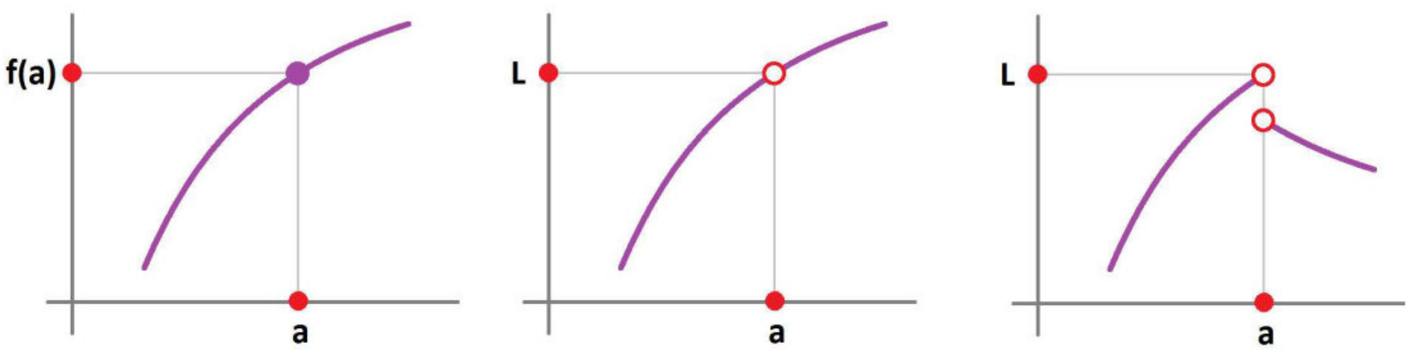


2.3. Limits of functions: small scale trends

The subject of our study is clear now: It is small scale behavior of functions. We ask, what features of the graph won't disappear no matter how much we zoom in?

One of the most crucial properties of a function is the integrity of its graph: *Is there a break or a cut?*

In addition, in order to study motion, we typically assume that to get from point A to point B, we have to visit every location between A and B:



If there is a jump in the graph of the function, it can't represent motion!

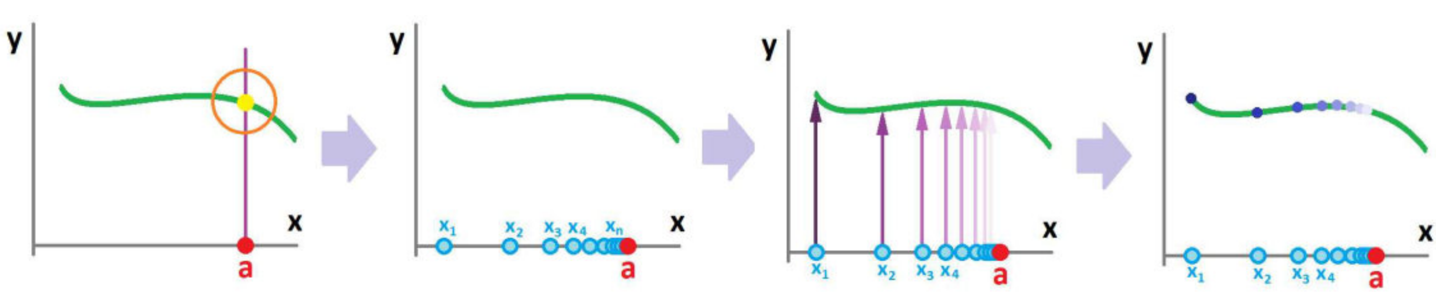
Furthermore, if it represents a transformation, there is tearing.

Thus, we want to understand what is happening to $y = f(x)$ when x is in the vicinity of a chosen point $x = a$.

The main tool with which we choose to test the behavior of a function around a point will be:

- *a sequence converging to the point,* $x_n \to a$, but never equal to it, $a_n \neq a$.

We then consider the effect of f on this sequence. First, we "lift" the points of the sequence from the x-axis to the graph of the function:



We then look for a possible long-term pattern of behavior of this new sequence:

▶ Do these points accumulate to another point on the graph?

Since we already know that the x-coordinates of these points do accumulate (to $x = a$), the question becomes: What is happening to the y-coordinates?

Example 2.3.1: three functions – three behaviors

Let's study these three functions around the point $a = 0$:

$$f(x) = \sin x, \quad g(x) = \frac{1}{x}, \quad h(x) = \sin(1/x)\,.$$

The reciprocal sequence is an appropriate choice for a test sequence:

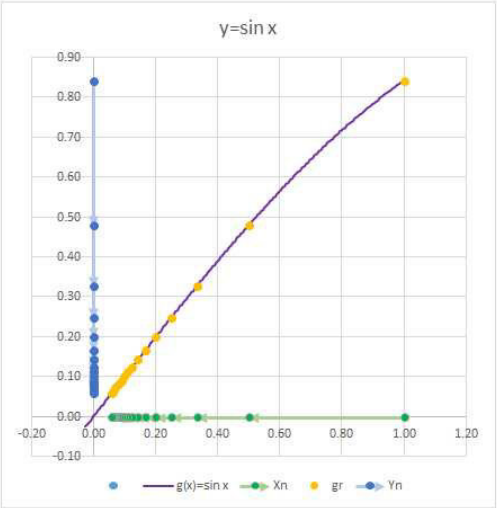
$$x_n = \frac{1}{n} \to 0\,.$$

The composition of a function and a sequence of x's gives us a new sequence, a sequence of y's. We have three, one for each function:

$$a_n = \cos(1/n), \quad b_n = \frac{1}{1/n}, \quad c_n = \sin\left(\frac{1}{1/n}\right)\,.$$

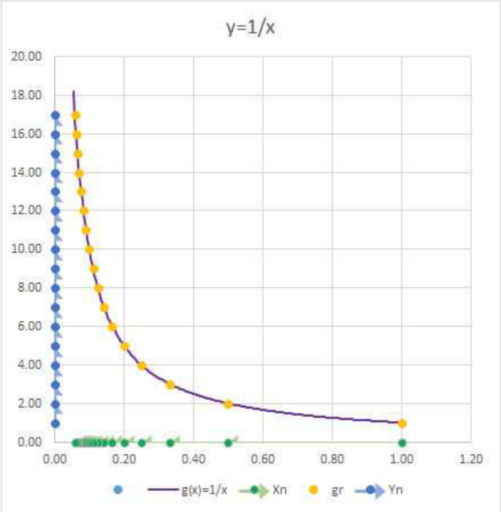
This is how the construction of these sequences can be visualized. We proceed from the x-axis to the graph to the y-axis as follows: We plot x_n on the x-axis, then (x_n, y_n) on the graph, and y_n on the y-axis.

Here is $f(x) = \sin x$:



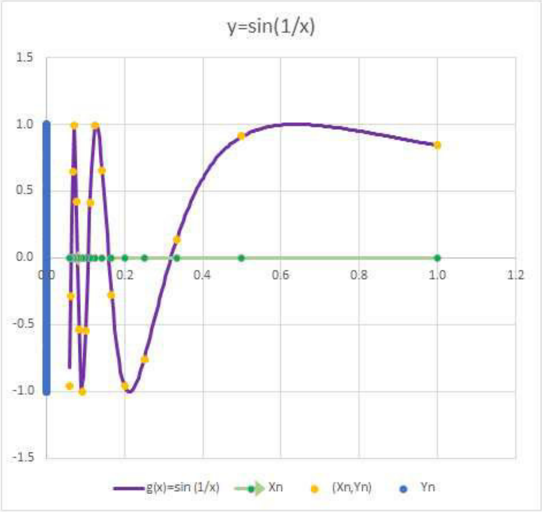
We see that the sequences x_n and y_n look very similar.

Here is $g(x) = 1/x$:



We see that x_n converges to 0 but y_n goes to infinity.

Here is $h(x) = \sin(1/x)$:



We see that x_n converges to 0 but y_n looks randomly spread around the interval $[-1, 1]$.

And this is what we discover about these new sequences when we explore them numerically:

$$a_n = \sin(1/n) \to 0, \quad b_n = n \to \infty, \quad c_n = \sin(n) \ \text{ no limit.}$$

Thus, the initial idea of how to find what is happening to $y = f(x)$ as x is approaching a is to pick a sequence that approaches a, i.e., $x_n \to a$. Then we evaluate this limit of a new sequence that comes from substitution:

$$\lim_{n\to\infty} f(x_n) = ?$$

If we think of the sequence x_n as a function, then we should interpret this substitution,

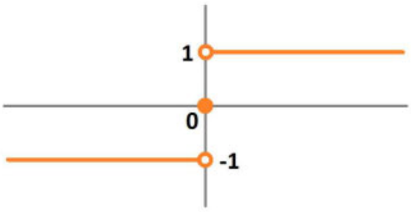
$$y_n = f(x_n)\,,$$

as the *composition*, as discussed in Chapter 1.

A single sequence might not be enough though!

Example 2.3.2: sign function

Just observe the following failure of a single sequence to tell us what is going on with the function sign around 0.



We try $x_n = -1/n$ and $x_n = 1/n$:

$$\lim_{n\to\infty} \operatorname{sign}(-1/n) = -1, \quad \text{but} \quad \lim_{n\to\infty} \operatorname{sign}(1/n) = 1\,,$$

as we approach 0 from one direction at a time. The limits don't match! In fact, the result suggests that the limit of $\operatorname{sign}(x)$ simply doesn't exist at this point. There is no trend!

Exercise 2.3.3

What test sequence should we choose to prove the discontinuity of this function?

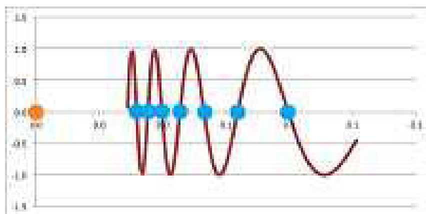
Example 2.3.4: sine of reciprocal

Another failure is of a single sequence test is as follows:

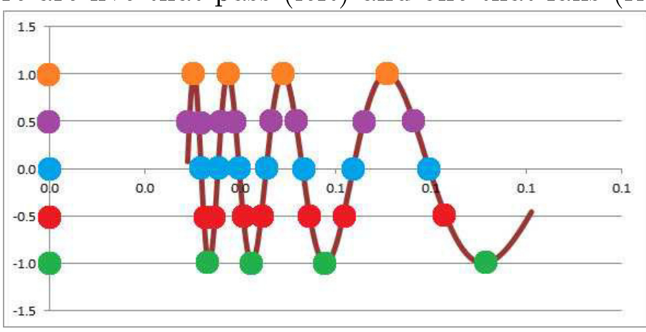
$$\lim_{n\to\infty} \sin\left(\frac{1}{1/n}\right) = \lim_{n\to\infty} \sin n \ \text{ doesn't exist,}$$

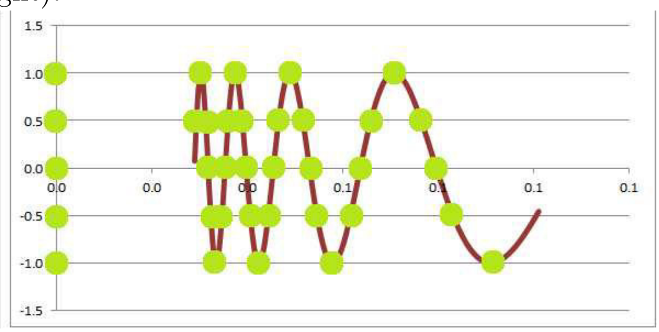
but

$$\lim_{n\to\infty} \sin\frac{1}{\pi n} = \lim_{n\to\infty} \sin(\pi n) = \lim_{n\to\infty} 0 = 0 \ \text{ does.}$$



Here are five that pass (left) and one that fails (right):





We try to substitute *all* sequences that converge to a and ensure they all induce the same behavior from f.

We finally summarize this idea here:

Definition 2.3.5: limit of function

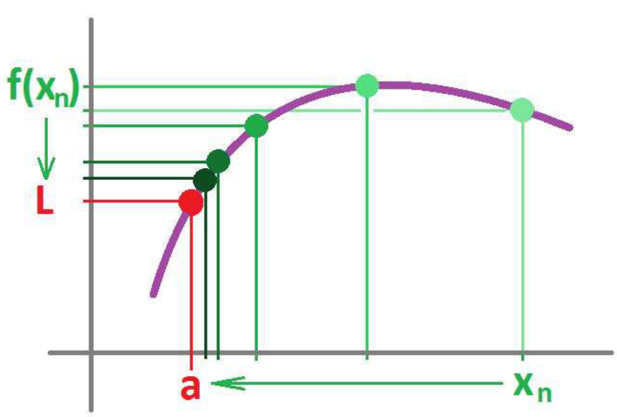
The *limit of a function* f at a point $x = a$ is defined to be the limit of its composition with any sequence x_n within the domain of f excluding a that converges to a,

$$x_n \to a \ \text{ with } x_n \neq a\,,$$

when all these limits exist, finite or infinite, and are equal to each other:

$$y_n = f(x_n) \to L\,.$$

As you can see, even though we set the issue of continuity aside for a moment, the method remains the same: We test the function with sequences. If a sequence of x's converges, does the corresponding sequence of y's converge too? That's a test:



But they all have to pass it. And the results have to be the same!

Example 2.3.6: "for any ... there is"

We shouldn't confuse the two familiar constructs in the last part of the definition. It is:

- `FOR ANY` sequence...

and not

- `THERE IS` a sequence...

The examples above illustrate the difference.

The development, the notation, and the terminology of the limits of functions follow that for sequences. Furthermore, the complexity of the definitions for the former is diminished because we rely on those for the latter.

We need to justify "the" in "the limit". Indeed, a limit is defined as a number that satisfies a certain property and there is no reason to assume that there must be only one.

Theorem 2.3.7: Uniqueness of Limit of Function

If we have two limits of the same function at the same point, they are equal.

Proof.

It follows from the uniqueness of the limit of a sequence presented in the last chapter.

We use the following notation for the limit of a function:

Limit of function

$$f(x)\to L \quad \text{as } x\to a$$

and

$$\lim_{x\to a} f(x)=L\,.$$

It reads "the limit of $f(x)$ as x is approaching a is L".

Warning!

The use of "$f(x)\to L$ as $x\to a$" might mislead one into thinking that the two parts are identical, while in fact $f(x)$ *can* be equal to L but x *cannot* be equal to a. Some like to use the following:

$$f(x)\to L \quad \text{as } x\to a^{\neq}\,.$$

We use a similar notation to describe the special kind of divergent behavior (infinity):

Infinite limit of function

$$f(x)\to \pm\infty \quad \text{as } x\to a$$

or

$$\lim_{x\to a} f(x)=\pm\infty\,.$$

It reads "the limit of $f(x)$ as x is approaching a is infinite".

We restate the theorem:

$$\boxed{L = \lim_{x\to a} f(x) \quad \texttt{AND} \quad M = \lim_{x\to a} f(x) \implies L = M}$$

The third possibility is that there is no trend. Since the definition of convergence starts with "`THERE IS` a number a...", we can say that otherwise there is no limit:

The limit does not exist

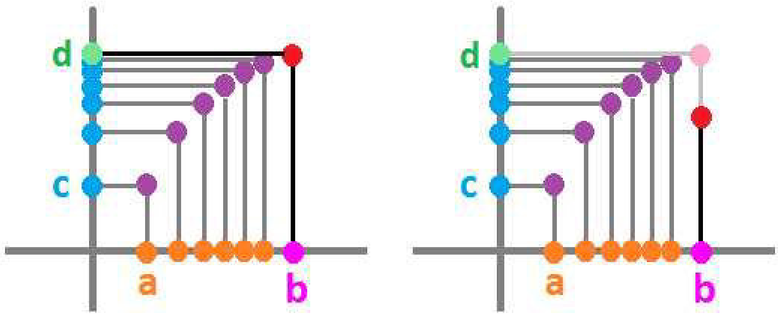
$$\lim_{x\to a} f(x), \quad \text{no limit}$$

or

$$\lim_{x\to a} f(x) \ \text{ DNE.}$$

It reads "the limit of $f(x)$ as x is approaching a does not exist".

Below, the limit does exist; it's just not equal to the value of the function:



The function is discontinuous! We have a new way to express continuity:

Theorem 2.3.8: Continuity via Limit

A function f is continuous at $x = a$ if and only if f is defined at a, the limit at a exists and is equal to the value of the function at a:

$$\boxed{\lim_{x\to a} f(x) = f(a)}$$

Example 2.3.9: limits from continuity

If the function is known to be continuous, the computation of the limit is trivial:

$$\begin{array}{lll}\lim_{x\to 3} 3x^3 + 2x^2 + 1 & = 3x^3 + 2x^2 + 1\Big|_{x=3} & = 3\cdot 3^3 + 2\cdot 3^2 + 1 \\ \lim_{x\to\pi} e^{x^2} & = e^{x^2}\Big|_{x=\pi} & = e^{\pi^2} \\ \lim_{x\to -1} \sin x & = \sin x\Big|_{x=-1} & = \sin(-1)\end{array}$$

Example 2.3.10: three test sequences

We will use these three "test sequences":

$$x_n = \frac{1}{n}, \quad y_n = -\frac{1}{n}, \quad z_n = \frac{(-1)^n}{n}\,.$$

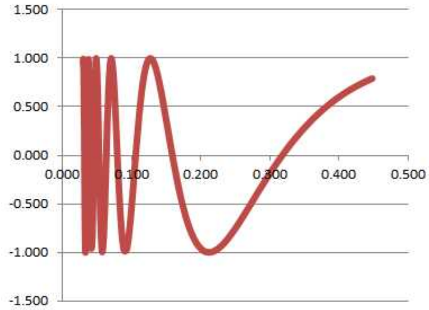
First, we take $f(x) = x^2$ at $a = 0$. The limits are the same:

$$\lim_{x\to a} f(x_n) = \lim_{x\to a} f(y_n) = \lim_{x\to a} f(z_n) = \lim_{x\to 0} \frac{1}{n^2} = 0\,.$$

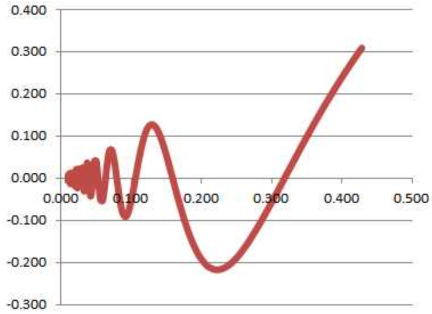
As we have seen, this limit does not exist:

$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right) \text{ , no limit.}$$

In fact, the values of this function start to fill the whole interval $[-1, 1]$ as we approach 0:



However, if we multiply this expression by x, the swings will start to diminish on the way to 0:



We have a limit:

$$\lim_{x\to 0} x \sin\left(\frac{1}{x}\right) = 0\,.$$

Let's use the alternating reciprocal sequence for $y = \operatorname{sign}(x)$:

$$x_n = (-1)^n \frac{1}{n}\,.$$

Then,

$$\operatorname{sign}(x_n) = \begin{cases} 1 & \text{if } x \text{ is even,} \\ -1 & \text{if } x \text{ is odd.} \end{cases}$$

This sequence is *divergent*. Therefore, the requirement of the definition fails, and $\lim_{x\to 0} \operatorname{sign}(x)$ doesn't exist.

Another way to come to this conclusion is to concentrate on one side at a time:

$$\begin{aligned} \lim_{n\to\infty} \operatorname{sign}\left(-\frac{1}{n}\right) &= \lim_{n\to\infty} -1 &&= -1\,, \\ \lim_{n\to\infty} \operatorname{sign}\left(\frac{1}{n}\right) &= \lim_{n\to\infty} 1 &&= 1\,. \end{aligned}$$

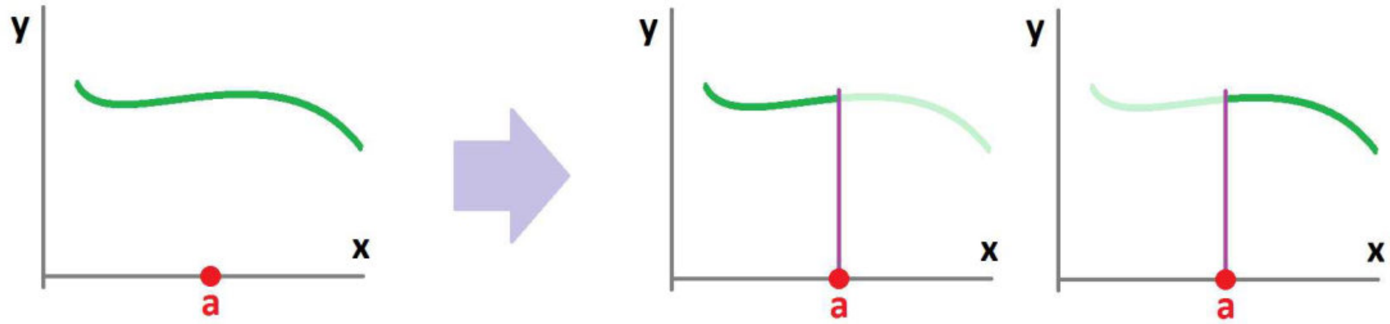
The two limits are *different*, the requirement fails, and that's why $\lim_{x\to 0} \operatorname{sign}(x)$ doesn't exist.

So, the behavior of $y = \operatorname{sign}(x)$ on the left and on the right, when considered separately, is very regular.

Indeed, we can choose any sequences and, as long as they stay on one side of 0, we have the same conclusion:

- If $x_n \to 0$ and $x_n < 0$ for all n, then $\lim\limits_{n\to\infty} \operatorname{sign}(x_n) = -1$.
- If $x_n \to 0$ and $x_n > 0$ for all n, then $\lim\limits_{n\to\infty} \operatorname{sign}(x_n) = 1$.

To take advantage of this insight, we can imagine that either the part of the graph of a function to the left of $x = a$ disappears or the part to the right does:



We make the idea precise below:

Definition 2.3.11: limit from the left and right

- The *limit from the left* of a function f at a point $x = a$ is defined to be the same limit, $\lim\limits_{n\to\infty} f(x_n)$, but only limited to the sequences x_n with
$$x_n \to a \ \text{ as } n \to \infty, \quad \text{and} \quad x_n < a \ \text{ for all } n\,,$$
when all these limits exist and are equal to each other.
- The *limit from the right* of a function f at a point $x = a$ is defined to be the same limit, $\lim\limits_{n\to\infty} f(x_n)$, but only limited to the sequences x_n with
$$x_n \to a \ \text{ as } n \to \infty, \quad \text{and} \quad x_n > a \ \text{ for all } n\,,$$
when all these limits exist and are equal to each other.

Otherwise, we say that *the limit from the right or from the left does not exist.* Collectively, the two are called *one-sided limits*.

In other words, one-sided limits are the limits of the function *restricted* to an interval to the left or to the right of the point:

Limit from the left: $\lim\limits_{x\to a^-} f(x)\Big|_{x<a}$ $\qquad - - - \circ$

Limit from the right: $\lim\limits_{x\to a^+} f(x)\Big|_{x>a}$ $\qquad \circ - - -$

The following notation is used:

One-sided limits

Left: $\lim\limits_{x\to a^-} f(x)$

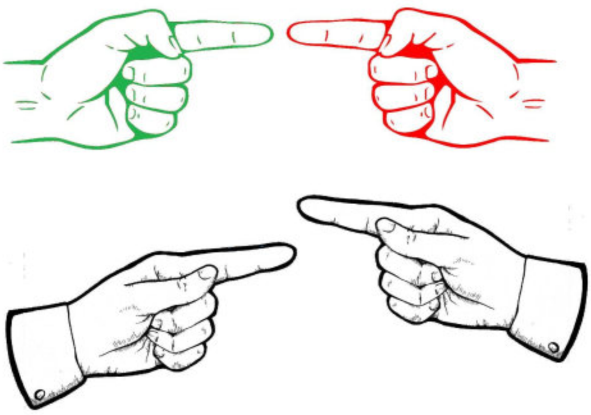
Right: $\lim\limits_{x\to a^+} f(x)$

- The notation a^- suggests that we only consider numbers of the type $a - \varepsilon$ with $\varepsilon > 0$.
- The notation a^+ suggests that we only consider numbers of the type $a + \varepsilon$ with $\varepsilon > 0$.

Warning!

Some sources also use $\lim\limits_{x\nearrow a} f(x)$ and $\lim\limits_{x\searrow a} f(x)$.

These are the "one-sided" limits. For the original, "two-sided" limit, the question becomes, do the two – left and right – limits match?



Theorem 2.3.12: Limit in Terms of One-Sided Limits

The limit of a function exists `IF AND ONLY IF` *the limits from the left and from the right of the point both exist and are equal to each other.*

In other words, we have:

$$\lim_{x\to a} f(x) \text{ exists} \iff \lim_{x\to a^-} f(x) = \lim_{x\to a^+} f(x)\,.$$

In that case, we have:

$$\lim_{x\to a} f(x) = \lim_{x\to a^+} f(x) = \lim_{x\to a^-} f(x)\,.$$

Proof.

The existence of this limit means that $\lim\limits_{n\to\infty} f(x_n)$ exists for any sequence $x_n \to a$ and is the same. In particular, this is true for the sequences limited to the ones all larger than a and all smaller than a. The proof of the converse is omitted.

Example 2.3.13: piecewise-defined functions

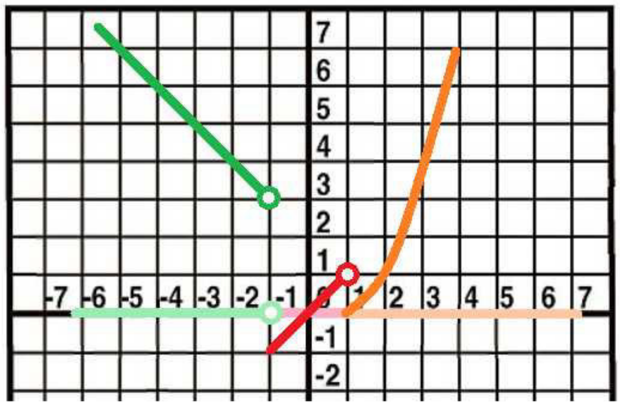
Let's plot the graph of this function:

$$f(x) = \begin{cases} 2-x & \text{if } x < -1\,, \\ x & \text{if } -1 < x < 1\,, \\ (x-1)^2 & \text{if } x > 1\,. \end{cases}$$

The limit $\lim\limits_{x\to a} f(x)$ exists for all a except $a = -1, 1$. While computing the limits, we pick the right formula from the three:

$$\begin{aligned} \lim_{x\to -1^-} f(x) &= \lim_{x\to -1^-} (2-x) &&= (2-(-1)) = 3 & \qquad \lim_{x\to -1^+} f(x) &= \lim_{x\to -1^+} x &&= -1 \\ \lim_{x\to 1^-} f(x) &= \lim_{x\to -1^-} x &&= 1 & \lim_{x\to 1^+} f(x) &= \lim_{x\to 1^+} (1-x)^2 &&= (1-1)^2 = 0 \end{aligned}$$

So, this is the graph made of three branches, one for each formula:



This example was about interpreting the graph in terms of limits. Now, the other way around.

Example 2.3.14: from limits to graphs

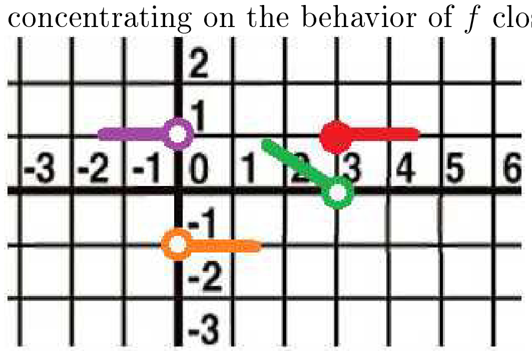
Given this information about f, plot its graph:

$$\begin{array}{llllll} \lim_{x\to 0^-} f(x) & = 1 & \lim_{x\to 0^+} f(x) & = -1 & f(0) & \text{undefined} \\ \lim_{x\to 3^-} f(x) & = 0 & \lim_{x\to 3^+} f(x) & = 1 & f(3) & = 1 \end{array}$$

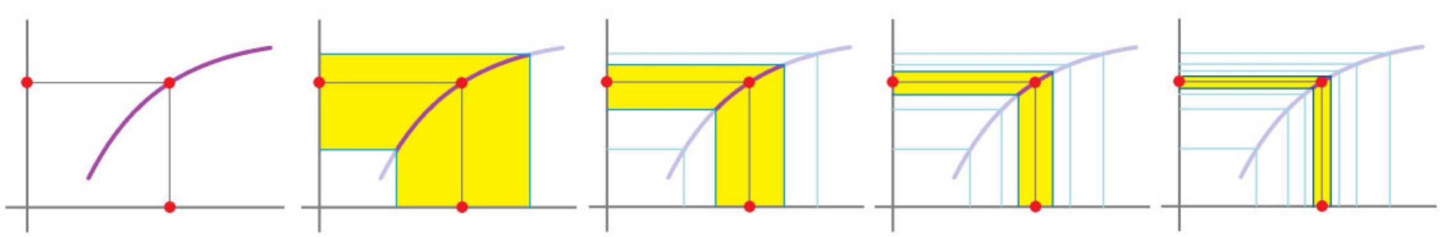
We rewrite the one-sided limits:

- As $x \to 0^-$, we have: $y \to 1$.
- As $x \to 0^+$, we have: $y \to -1$.
- As $x \to 3^-$, we have: $y \to 0$.
- As $x \to 3^+$, we have: $y \to 1$.

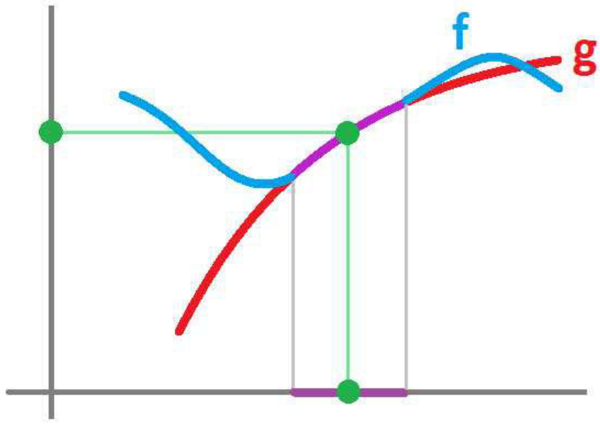
Then we plot the results below concentrating on the behavior of f close to these points:



If we consider two sequences converging to a – from the left and from the right – we see a sequence of *intervals* shrinking towards a:



As we know, the limit of a sequence is fully determined by its *tail*. Therefore, the limit of a function at a is fully determined by the tail of the sequence $f(x_n)$ when $x_n \to a$. If we can only see the behavior of f in the vicinity – no matter how small – of a, we still know the value (and the existence) of the limit $\lim\limits_{x\to a} f(x)$:



This is our conclusion:

Theorem 2.3.15: Limits Are Local

The limit of a function is determined by its values in the vicinity of the point.

In other words, if two functions f and g are equal over an open interval that contains the point a:

$$f(x) = g(x) \quad \text{for all } a - \varepsilon < x < a + \varepsilon,\ x \neq a\,,$$

for some $\varepsilon > 0$, then their limits at a coincide too:

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x)\,,$$

unless neither exists.

Exercise 2.3.16

Restate the theorem in terms of restrictions of functions.

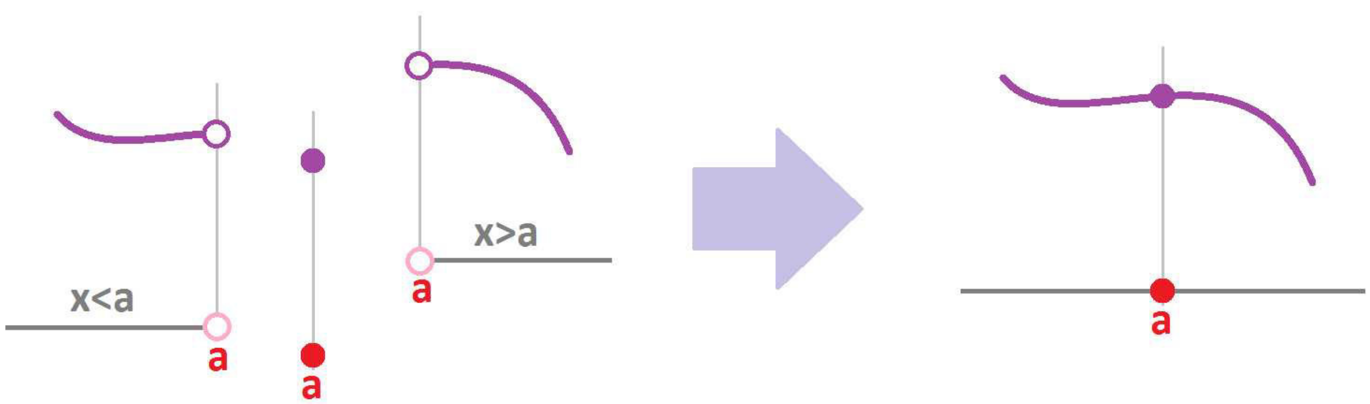
Exercise 2.3.17

State an analog of the theorem for one-sided limits.

Recall that for a function f and a point a, where f is defined, the graph of f consists of three parts:

1. the part of the graph of f with $x < a$,
2. the part of the graph of f with $x = a$ (one point), and
3. the part of the graph of f with $x > a$.

For this function to be continuous, these three parts (the two pieces of the rope and a drop of glue) have to fit together:



We put this idea in the form of a theorem that relies on the concept of one-sided limit:

Theorem 2.3.18: Continuity via One-Sided Continuity

A function f is continuous at $x = a$ if and only if f is defined at a, the two one-sided limits exist and both equal to the value of the function at a.

In other words, we have:

$$\lim_{x\to a^-} f(x) = f(a) = \lim_{x\to a^+} f(x)\,.$$

Exercise 2.3.19

Make a hand-drawn sketch of the graph of the function and evaluate its continuity:

$$f(x) = \begin{cases} -3 & \text{if } x < 0, \\ x^2 & \text{if } 0 \le x < 1, \\ x & \text{if } x > 1. \end{cases}$$

2.4. Limits and continuity under algebraic operations

The limit procedure for a given a is a special kind of function, a function whose input is a numerical *function*:

$$\text{function} \quad \to \quad \boxed{\lim_{x\to a}} \quad \to \quad \text{a number}$$

Exercise 2.4.1

What can you say about the domain of this special function?

We will use the algebraic properties of the limits of sequences to prove virtually identical facts about limits of functions.

Let's review the main algebraic properties:

Theorem 2.4.2: Algebra of Limits of Sequences

Suppose $a_n \to a$ and $b_n \to b$. Then we have:

$$\boxed{\begin{array}{lll|lll} \textit{SR:} & a_n + b_n \to a + b & & \textit{CMR:} & c \cdot a_n \to c \cdot a & \textit{for any real } c \\ \textit{PR:} & a_n \cdot b_n \to a \cdot b & & \textit{QR:} & a_n / b_n \to a/b & \textit{provided } b \ne 0 \end{array}}$$

Presented verbally, these rules have these abbreviated versions:

- *Sum Rule*: The limit of the sum is the sum of the limits.
- *Constant Multiple Rule*: The limit of a constant multiple is the constant multiple of the limits.
- *Product Rule*: The limit of the product is the product of the limits.
- *Quotient Rule*: The limit of the quotient is the quotient of the limits as long as the limit of the denominator isn't zero.

Each property is matched by its analog for functions. Furthermore, there are analogs for the one-sided limits. In fact, we will see analogs of some of these rules for numerous new concepts in the forthcoming chapters:

Theorem 2.4.3: Algebra of Limits of Functions

Suppose $f(x) \to F$ *and* $g(x) \to G$ *as* $x \to a$ (or $x \to a^-$ or $x \to a^+$). *Then*

SR:	$f(x) + g(x) \to F + G$	*CMR:*	$c \cdot f(x) \to c \cdot F$	*for any real* c
PR:	$f(x) \cdot g(x) \to F \cdot G$	*QR:*	$f(x)/g(x) \to F/G$	*provided* $G \neq 0$

Let's consider them all starting with the following:

Theorem 2.4.4: Sum Rule of Limits of Functions

If the limits at a *of functions* $f(x), g(x)$ *exist, then so does that of their sum,* $f(x) + g(x)$.

Furthermore, the limit of the sum is equal to the sum of the limits:

$$\lim_{x\to a} \big(f(x) + g(x)\big) = \lim_{x\to a} f(x) + \lim_{x\to a} g(x)\,.$$

Proof.

For any sequence $x_n \to a$, we have by Sum Rule for Sequences:

$$\lim_{x\to a}(f(x) + g(x)) = \lim_{n\to\infty}(f(x_n) + g(x_n)) = \lim_{n\to\infty} f(x_n) + \lim_{n\to\infty} g(x_n)\,.$$

In the case of infinite limits, we follow the rules of the algebra of infinities as in Chapter 1:

number	$+$	$(+\infty)$	$= +\infty$
number	$+$	$(-\infty)$	$= -\infty$
$+\infty$	$+$	$(+\infty)$	$= +\infty$
$-\infty$	$+$	$(-\infty)$	$= -\infty$

Warning!

Indeterminate expressions aren't here: ∞/∞, $\infty - \infty$, $0/0$, etc.

The proofs of the rest of the properties are identical. The results are presented below:

Theorem 2.4.5: Constant Multiple Rule of Limits of Functions

If the limit at a *of function* $f(x)$ *exists, then so does that of its multiple,* $cf(x)$.

Furthermore, the limit of the multiple is equal to the multiple of the limit:

$$\lim_{x\to a} cf(x) = c \cdot \lim_{x\to a} f(x)\,.$$

Theorem 2.4.6: Product Rule of Limits for Functions

If the limits at a of functions $f(x), g(x)$ exist, then so does that of their product, $f(x) \cdot g(x)$.

Furthermore, the limit of the product is equal to the product of the limits:

$$\lim_{x\to a}\big(f(x)\cdot g(x)\big) = \big(\lim_{x\to a} f(x)\big)\cdot\big(\lim_{x\to a} g(x)\big)\,.$$

Theorem 2.4.7: Quotient Rule of Limits of Functions

If the limits at a of functions $f(x), g(x)$ exist, then so does that of their ratio, $f(x)/g(x)$, provided $\lim\limits_{x\to a} g(x) \neq 0$.

Furthermore, the limit of the ratio is equal to the ratio of the limits:

$$\lim_{x\to a}\left(\frac{f(x)}{g(x)}\right) = \frac{\lim\limits_{x\to a} f(x)}{\lim\limits_{x\to a} g(x)}\,.$$

We can say that the limit sign is "distributed" over these algebraic operations.

These results are known as the *Limits Rules*.

The main building blocks are these two functions – the constant function and the identity function – with simple limits:

Theorem 2.4.8: Limit of Constant Function

For any real c, the following limit exists at any point a:

$$\lim_{x\to a} c = c\,.$$

Proof.

For any sequence $x_n \to a$, we have:

$$\lim_{x\to a} c = \lim_{n\to\infty} c = c\,.$$

Let's set $g(x) = c$ in *Product Rule* and use the last theorem, then

$$\lim_{x\to a} cf(x) = \lim_{x\to a}(f(x)\cdot g(x)) = c\cdot(\lim_{x\to a} g(x))\,.$$

Then the *Constant Multiple Rule* follows. Even though the *Constant Multiple Rule* is absorbed into *Product Rule*, the former is simpler and easier to use:

Theorem 2.4.9: Limit of Identity Function

The following limit exists at any point a:

$$\lim_{x\to a} x = a\,.$$

Proof.

For any sequence $x_n \to a$, we have:

$$\lim_{x\to a} x = \lim_{n\to\infty} x_n = a\,.$$

Any polynomial can be built from x and constants by multiplication and addition. Therefore, the first five theorems allow us to compute the limits of all polynomials.

Example 2.4.10: polynomials

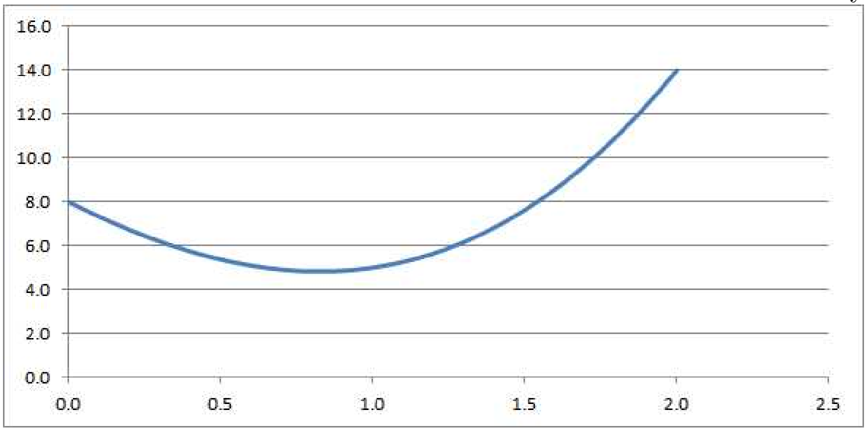
Let
$$f(x) = x^3 + 3x^2 - 7x + 8\,.$$
What is its limit as $x \to 1$? The computation is straightforward, but every step has to be justified with the rules above.

To understand which rules to apply *first*, observe what the *last* operation is; it's addition. We use the *Sum Rule*, subject to justification:

$$\begin{aligned}
\lim_{x\to 1} f(x) &= \lim_{x\to 1}(x^3 + 3x^2 - 7x + 8) && \text{Continue with SR.}\\
&= \lim_{x\to 1} x^3 + \lim_{x\to 1} 3x^2 - \lim_{x\to 1} 7x + \lim_{x\to 1} 8 && \text{Continue with}\\
&\qquad \text{PR,} \qquad\quad \text{CMR,} \qquad \text{CMR,} \qquad \text{CR.} &&\\
&= \lim_{x\to 1} x \cdot \lim_{x\to 1} x^2 + 3\lim_{x\to 1} x^2 - 7\lim_{x\to 1} x + 8 && \text{Continue with CR.}\\
&= 1 \cdot \lim_{x\to 1} x^2 + 3\lim_{x\to 1} x^2 - 7\cdot 1 + 8 && \text{Continue with PR and CR.}\\
&= 1\cdot 1 + 3\cdot 1 - 7 + 8\\
&= 5\,.
\end{aligned}$$

With this complex argument, it is easy to miss the simple fact that the limit of this function happens to be equal to its *value*:
$$\lim_{x\to 1} f(x) = \lim_{x\to 1}(x^3 + 3x^2 - 7x + 8) = x^3 + 3x^2 - 7x + 8\Big|_{x=1} = 1^3 + 3\cdot 1^2 - 7\cdot 1 + 8 = 5\,.$$

We know this is true because the function is continuous! The idea is confirmed by the plot:



Let's re-state the definition of continuity:

Definition 2.4.11: continuity

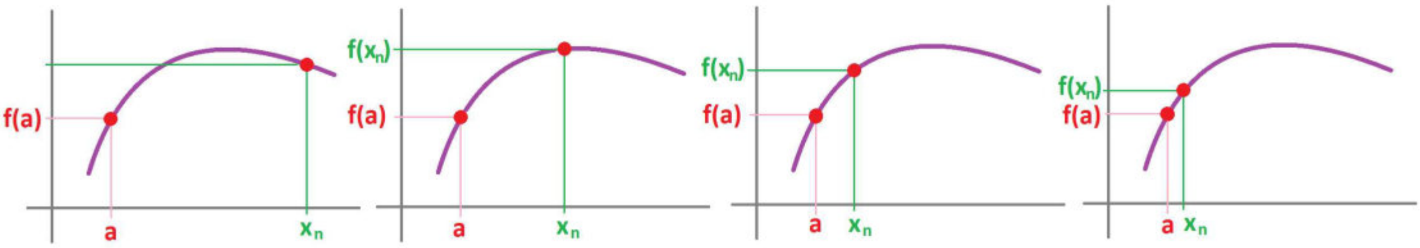
A function f is called *continuous at point* a if

- $f(x)$ is defined at $x = a$,
- the limit of f exists at a, and
- the two are equal to each other:

$$\lim_{x\to a} f(x) = f(a)\,.$$

Thus, the limits of continuous functions can be found by *substitution*.

Equivalently, a function f is continuous at a if
$$\lim_{n\to\infty} f(x_n) = f(a)\,,$$
for any sequence $x_n \to a$:



As we shall see, all polynomials are continuous. Now an example of a rational function:

Example 2.4.12: rational functions

Let's find the limit at 2 of

$$f(x)=\frac{x+1}{x-1}\,.$$

Again, we look at the *last* operation of the function. It is division, so we use *Quotient Rule* first:

$$\begin{aligned}\lim_{x\to 2} f(x) &= \lim_{x\to 2}\frac{x+1}{x-1} && \text{We now justify QR by observing that}\\ & && \lim_{x\to 2}(x-1)=1\neq 0\,.\\ &= \frac{\lim_{x\to 2}(x+1)}{\lim_{x\to 2}(x-1)}\\ &= \frac{3}{1}\\ &= 3\,.\end{aligned}$$

Example 2.4.13: indeterminate expressions

Let's find the limit at 1 of the function

$$f(x)=\frac{x^2-1}{x-1}\,.$$

Since the last operation is division, we are supposed to use the *Quotient Rule* first. However, the limit of the denominator is 0:

$$\lim_{x\to 1}(x-1)=0\,.$$

Then, the *Quotient Rule* is inapplicable. But then all other rules of limits are also inapplicable!

A closer look reveals that things are even worse; both the numerator and the denominator go to 0 as x goes to 1. An attempt to apply the *Quotient Rule* – over these objections – would result in an *indeterminate expression*:

$$\frac{x^2-1}{x-1}\xrightarrow{???}\frac{0}{0}\ \text{ as } x\to 1\quad \boxed{\textbf{DEAD END}}$$

This doesn't mean that the limit doesn't exist; it doesn't mean anything! The only conclusion we can draw from this is that we have done something we shouldn't have. We go back and, instead of QR,

do some algebra.

We factor the numerator and then cancel the denominator (thereby circumventing the need for QR):

$$\begin{aligned}\lim_{x\to1} f(x) &= \lim_{x\to1}\frac{x^2-1}{x-1}\\ &= \lim_{x\to1}\frac{(x-1)(x+1)}{x-1} \quad \text{for} \quad x\neq 1\\ &= \lim_{x\to1}(x+1)\\ &= 2\,.\end{aligned}$$

The cancellation is justified by the fact that x tends to 1 but never reaches it.

Let's consider more examples of how trying to apply the laws of limits without verifying their conditions could lead to indeterminate expressions.

Example 2.4.14: indeterminate $0/0$

We choose a few algebraically trivial situations. For $x \to 0$ below, misapplications of the *Quotient Rule* lead to the same problem. When resolved, the answers *vary*:

function	wrong turn	outcome	redone with algebra	outcome
$\frac{x^2}{x}$	$\xrightarrow{???} \frac{0}{0}$	**DEAD END**	$\frac{x^2}{x} = x$	$\to 0$
$\frac{x}{x^2}$	$\xrightarrow{???} \frac{0}{0}$	**DEAD END**	$\frac{x}{x^2} = \frac{1}{x}$	$\to \infty$
$\frac{x}{x}$	$\xrightarrow{???} \frac{0}{0}$	**DEAD END**	$\frac{x}{x} = 1$	$\to 1$
$\frac{x\sin(1/x)}{x}$	$\xrightarrow{???} \frac{0}{0}$	**DEAD END**	$\frac{x\sin(1/x)}{x} = x\sin(1/x)$	DNE

So, *any* answer is possible when we face an indeterminate expression!

Example 2.4.15: indeterminate ∞/∞

Now, there are other kinds of indeterminate expressions. For $x \to 0$, a misapplication of the *Quotient Rule* leads to the following:

function	wrong turn	outcome	redone with algebra	outcome
$\frac{1/x}{1/x^2}$	$\xrightarrow{???} \frac{\infty}{\infty}$	**DEAD END**	$\frac{1/x}{1/x^2} = x$	$\to 0$
$\frac{1/x^2}{1/x}$	$\xrightarrow{???} \frac{\infty}{\infty}$	**DEAD END**	$\frac{1/x^2}{1/x} = \frac{1}{x}$	$\to \infty$

So, *any* answer is possible when we face an indeterminate expression!

Exercise 2.4.16

Show more examples of different outcomes for the indeterminate expression above.

Example 2.4.17: indeterminate $\infty - \infty$

Finally, we see how indeterminate expressions appear under the *Sum Rule* instead of the *Quotient Rule*. For $x \to 0$, we have the following:

function	wrong turn	outcome	redone with algebra	outcome
$(1/x+1)-1/x$	$\xrightarrow{???} \infty - \infty$	**DEAD END**	$(1/x+1)-1/x=1$	$\to 1$

Exercise 2.4.18

Show more examples of different outcomes for the indeterminate expression above.

Example 2.4.19: algebraic tricks

Compute

$$\lim_{x\to 0} \frac{\sqrt{x^2+9}-3}{x^2}\,.$$

If we mindlessly substitute $x=0$, we get $0/0$. What does it mean? It means:

DEAD END

STOP! Erase everything and do algebra.

The goal is to cancel the denominator. The trick is to multiply by the *conjugate*, $\sqrt{x^2+9}+3$, of the numerator in order to "rationalize" it. Then we have:

$$\begin{aligned}
\lim_{x\to 0} \frac{\sqrt{x^2+9}-3}{x^2} &= \lim_{x\to 0} \frac{(\sqrt{x^2+9}-3)\cdot(\sqrt{x^2+9}+3)}{x^2\cdot(\sqrt{x^2+9}+3)}\\
&= \lim_{x\to 0} \frac{(x^2+9)-3^2}{x^2(\sqrt{x^2+9}+3)}\\
&= \lim_{x\to 0} \frac{x^2}{x^2(\sqrt{x^2+9}+3)}\\
&= \lim_{x\to 0} \frac{1}{\sqrt{x^2+9}+3}\\
&= \frac{1}{\sqrt{0+9}+3}\\
&= \frac{1}{6}\,.
\end{aligned}$$

At the end, the *Quotient Rule* applies because the limit in the denominator exists and is not 0.

Example 2.4.20: two famous limits

We are in the same position (0/0 indeterminacy) with these two familiar limits:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1 \text{ and } \lim_{x\to 0} \frac{1-\cos x}{x} = 0\,.$$

They are resolved later with trigonometry. They would produce indeterminate expressions if we tried to apply the rules of limits without checking their conditions first.

Warning!

The answer to a limit problem can't be "It's indeterminate!".

Example 2.4.21: algebra with DNE

Is it possible that $\lim\limits_{x\to a}(f(x)+g(x))$ exists, even though $\lim\limits_{x\to a} f(x)$ and $\lim\limits_{x\to a} g(x)$ do not? In other words, can their addition, $f+g$, cancel their irregular behavior? Of course; just pick $g=-f$. Then $f+g=0$, so $\lim\limits_{x\to a}(f+g)=0$. For specific examples, we can take:

- $f(x)=\dfrac{1}{x}$ and $a=0$. The limit does not exist:
$$\lim_{x\to 0}\frac{1}{x}\,.$$
- Neither does this:
$$\lim_{x\to 0}\left(-\frac{1}{x}\right)\,.$$
- But this one does:
$$\lim_{x\to 0}\left[\frac{1}{x}+\left(-\frac{1}{x}\right)\right]=\lim_{x\to 0}0=0\,.$$

We combine limits that exists with those that don't below:

Theorem 2.4.22: Divergence under Addition

If $\lim\limits_{x\to a} f(x)$ exists but $\lim\limits_{x\to a} g(x)$ does not, then $\lim\limits_{x\to a}(f(x)+g(x))$ does not exist.

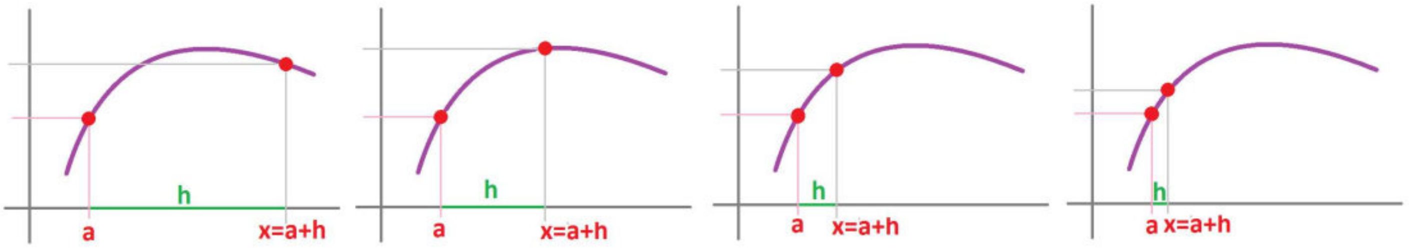
We take a new look at the behavior of x in the definition of limit. Instead of concentrating on

- ▶ how x is approaching a,

we can look at

- ▶ how far x is from a.

We consider the *increment*, i.e., the difference h between the two:



Thus, we have an equivalence:

$$\boxed{x\to a \iff h=x-a\to 0}$$

Then we rewrite the limit $\lim\limits_{x\to a} f(x)$ by substituting $h=x-a$, as follows:

Theorem 2.4.23: Alternative Formula for Limit of Function

The limit of a function f at a is equal to L if and only if

$$\boxed{\lim_{h\to 0} f(a+h)=L}$$

This will make our computations look different.

Example 2.4.24: easier cancellation

Let's use the theorem to take another approach to the following limit:

$$\begin{aligned}
\lim_{x\to 1}\frac{x^2-1}{x-1} &= \lim_{h\to 0}\frac{(1+h)^2-1}{(1+h)-1} && \text{According to the theorem.}\\
&= \lim_{h\to 0}\frac{1+2h+h^2-1}{1+h-1} && \text{We expand.}\\
&= \lim_{h\to 0}\frac{2h+h^2}{h} && \text{We simplify.}\\
&= \lim_{h\to 0}(2+h) && \text{We divide.}\\
&= 2+h\Big|_{h=0} && \text{We substitute.}\\
&= 2\,.
\end{aligned}$$

This time, we didn't have to do any factoring!

As we see, this substitution might make the algebra simpler.

Exercise 2.4.25

Use the theorem to compute:

$$\lim_{x\to 1}\frac{x^3-1}{x-1}\,.$$

Just as with sequences, we can represent the *Sum Rule* as a diagram:

$$\begin{tikzcd}
f,g \arrow[r, "\lim"] \arrow[d, "+"] \arrow[dr, phantom, "SR"] & F,G \arrow[d, "+"] \\
f+g \arrow[r, "\lim"] & \lim(f+g)=F+G
\end{tikzcd}$$

In the diagram, we start with a pair of functions at the top left and then we can proceed in two ways:

- Right: take the limit of either; then down: add the results.
- Down: add them; then right: take the limit of the result.

The result is the same! For the Product Rule and the Quotient Rule, we just replace "+" with "·" and "÷" respectively.

With the apparent abundance of discontinuous functions, now the good news:

▶ A typical function we encounter is continuous at every point of its domain.

Of course, it can't be continuous outside the domain. From all the functions we have seen so far, only a few piecewise-defined functions have been exceptions: the sign function and the integer value function.

We prove the continuity of a function by showing the fact that its limit is evaluated by substitution:

$$\boxed{\lim_{x\to a}F(x)=F(a) \quad \text{or} \quad F(x)\to F(a) \text{ as } x\to a}$$

Once it is proven, we turn this around and use this fact every time we need to compute a limit. The following theorem helps:

Theorem 2.4.26: Algebra and Continuity

Suppose f and g are continuous at $x = a$. Then so are the following functions:

SR:	$f+g$	*CMR:*	$c \cdot f$	*for any real c*
PR:	$f \cdot g$	*QR:*	f/g	*provided $g(a) \neq 0$*

Proof.

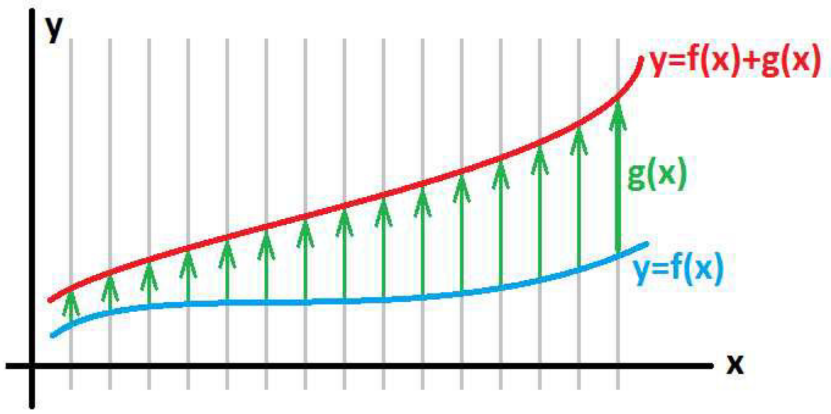
For the *Sum Rule*, we try to evaluate the limit of the sum at a as follows:

$$\begin{aligned}\lim_{x\to a}\left[(f+g)(x)\right] &= \lim_{x\to a}\left[f(x)+g(x)\right] && \text{We use SR next.}\\ &= \lim_{x\to a} f(x) + \lim_{x\to a} g(x) && \text{We use continuity next.}\\ &= f(a)+g(a)\\ &= (f+g)(a)\,.\end{aligned}$$

Therefore, the limit of $f+g$ is the value of the function; hence, it is continuous by the definition. Next, the *Constant Multiple Rule*, the *Product Rule*, and the *Quotient Rule* are proven with the *Constant Multiple Rule*, the *Product Rule*, and the *Quotient Rule* for limits, respectively.

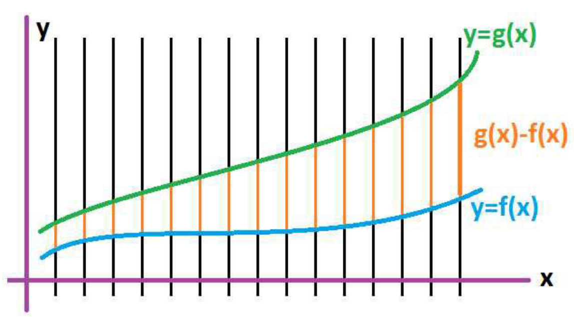
Let's review the geometric meaning of these rules.

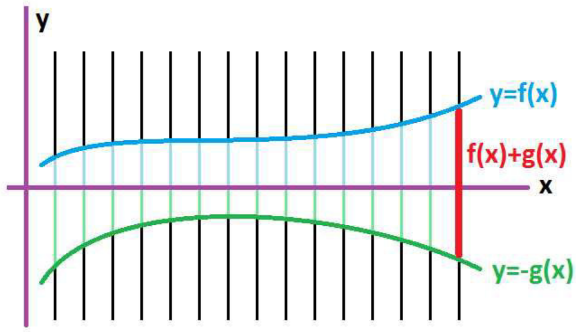
In the *Sum Rule*, g serves as a vertical "push" of the graph of f. The picture below is meant to illustrate that idea. There are ping-pong balls arranged in a curve, f, on the ground and there is also wind, g. Then, the wind, non-uniformly but continuously, blows them forward:



The ping-pong balls remain arranged in a curve, $f + g$.

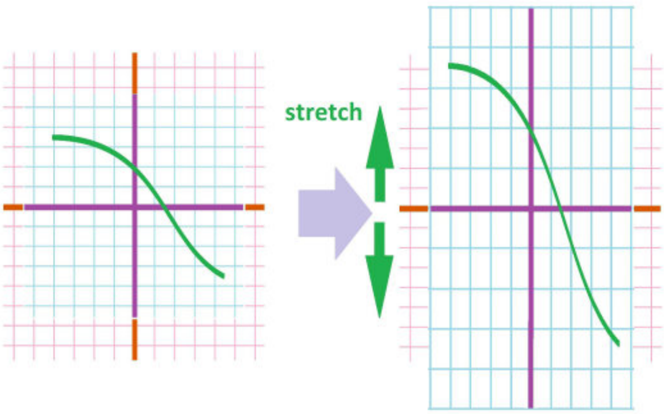
We can say that if the floor and the ceiling of a tunnel represented by f and g respectively are changing continuously, then so is its height, which is $g - f$ (left):



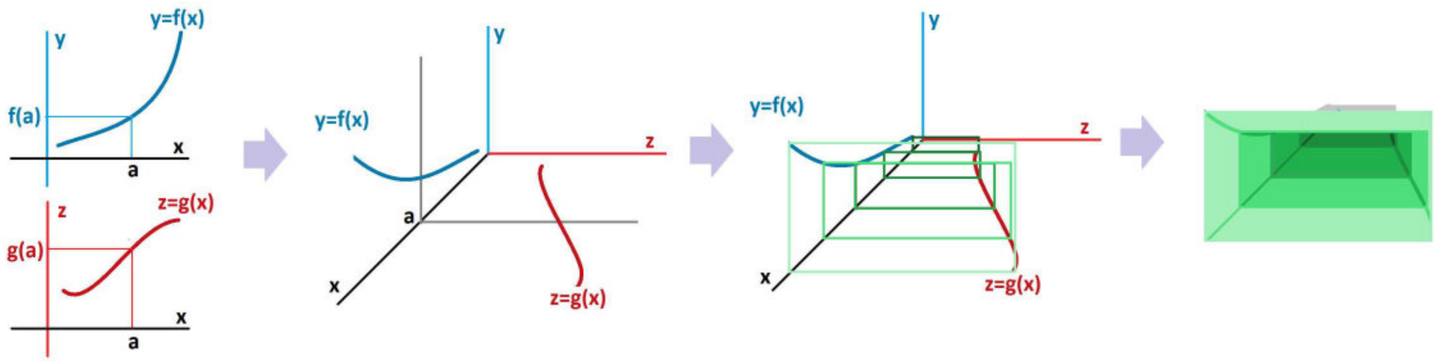


Or, if the floor and the ceiling, f and $-g$, of a tunnel are changing continuously, then so is its height, $g + f$ (right). Thus, if there are no bumps on the floor and no bumps on the ceiling, you won't (suddenly) bump your head as you walk.

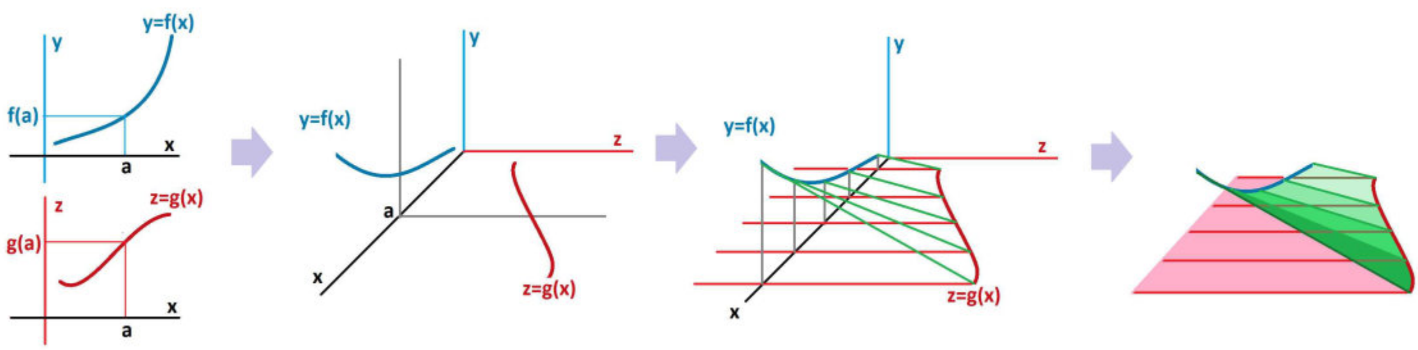
In the *Constant Multiple Rule*, the multiple c is the magnitude of a vertical stretch/shrink of the rubber sheet that has the graph of f drawn on it:



In the *Product Rule*, we say that if the width and the height, f and g, of a rectangle are changing continuously, then so is its area, $f \cdot g$:



In the *Quotient Rule*, we say that if the width and the height, f and g, of a triangle are changing continuously, then so is the tangent of its base angle, f/g:



The last one stands out because of its extra restriction that the denominator isn't zero – but only at the point a itself. For example, $\dfrac{1}{x-1}$ is continuous at $x=0$; the fact that it is undefined at 1 is irrelevant. On the other hand, division by 0 creates a hole in the domain; the function can't be continuous there anyway!

An abbreviated version of the theorem reads as follows:

Corollary 2.4.27: Continuity under Algebra

The sum, the difference, the product, and the ratio of two continuous functions is continuous (on its domain).

Exercise 2.4.28

Sketch the graphs of the polynomials with the following parameters:

	a	b	c
degree	1	3	4
the leading coefficient	-2	2	-1

To justify our conclusion about the *continuity of polynomials*, let's consider a general representation of an nth degree polynomial:

$$a_0 + a_1x + a_2x^2 + ... + a_{n-1}x + a_nx^n \,.$$

Then we follow the following sequence of conclusions:

$$\begin{array}{lccccccccccc}
\text{These are continuous:} & 1 & , & x & , & x & , & ... & & x & , & x\,. \\
\text{These are too by PR:} & 1 & , & x & , & x^2 & , & ... & & x^{n-1} & , & x^n\,. \\
\text{These too by CMR:} & a_0 & , & a_1x & , & a_2x^2 & , & ... & & a_{n-1}x^{n-1} & , & a_nx^n\,. \\
\text{This is continuous by SR:} & a_0 & + & a_1x & + & a_2x^2 & + & ... & + & a_{n-1}x^{n-1} & + & a_nx^n\,.
\end{array}$$

We also conclude that the graph of a polynomial consists of a single piece.

Example 2.4.29: limits by substitution

Evaluation of the limit of a polynomial is now elementary:

$$\lim_{x\to 22}\left(x^2 - 17x^2 + 7x - 2\right) = 22^2 - 17\cdot 22^2 + 7\cdot 22 - 2\,.$$

The following result is a very useful shortcut:

Theorem 2.4.30: Continuity of Polynomials and Rational Functions

- *Every polynomial is continuous at every point.*
- *Every rational function is continuous at every point where it is defined.*

Exercise 2.4.31

Restate as an implication and discuss the converse: "Every polynomial is a rational function".

Once the continuity of polynomials is established, the continuity of rational functions (away from the points where the denominator is zero) is proven from the *Quotient Rule*. We also conclude that the graph of a rational function consists of several pieces – one for each interval of its domain.

Example 2.4.32: rational

For example, the function

$$f(x) = \frac{1}{(x-1)^2(x-2)(x-3)}$$

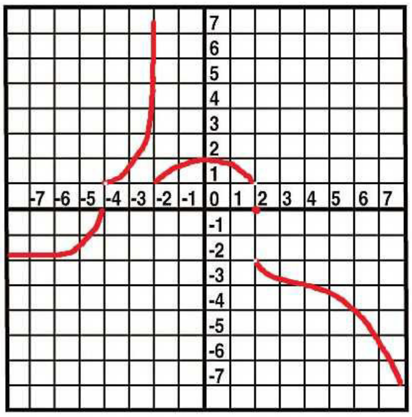
has three holes, $x = 1, 2, 3$, in its domain:

$$\begin{array}{ccccccc}
 & 1 & & 2 & & 3 & \\
----- & \circ & ----- & \circ & ----- & \circ & -----
\end{array}$$

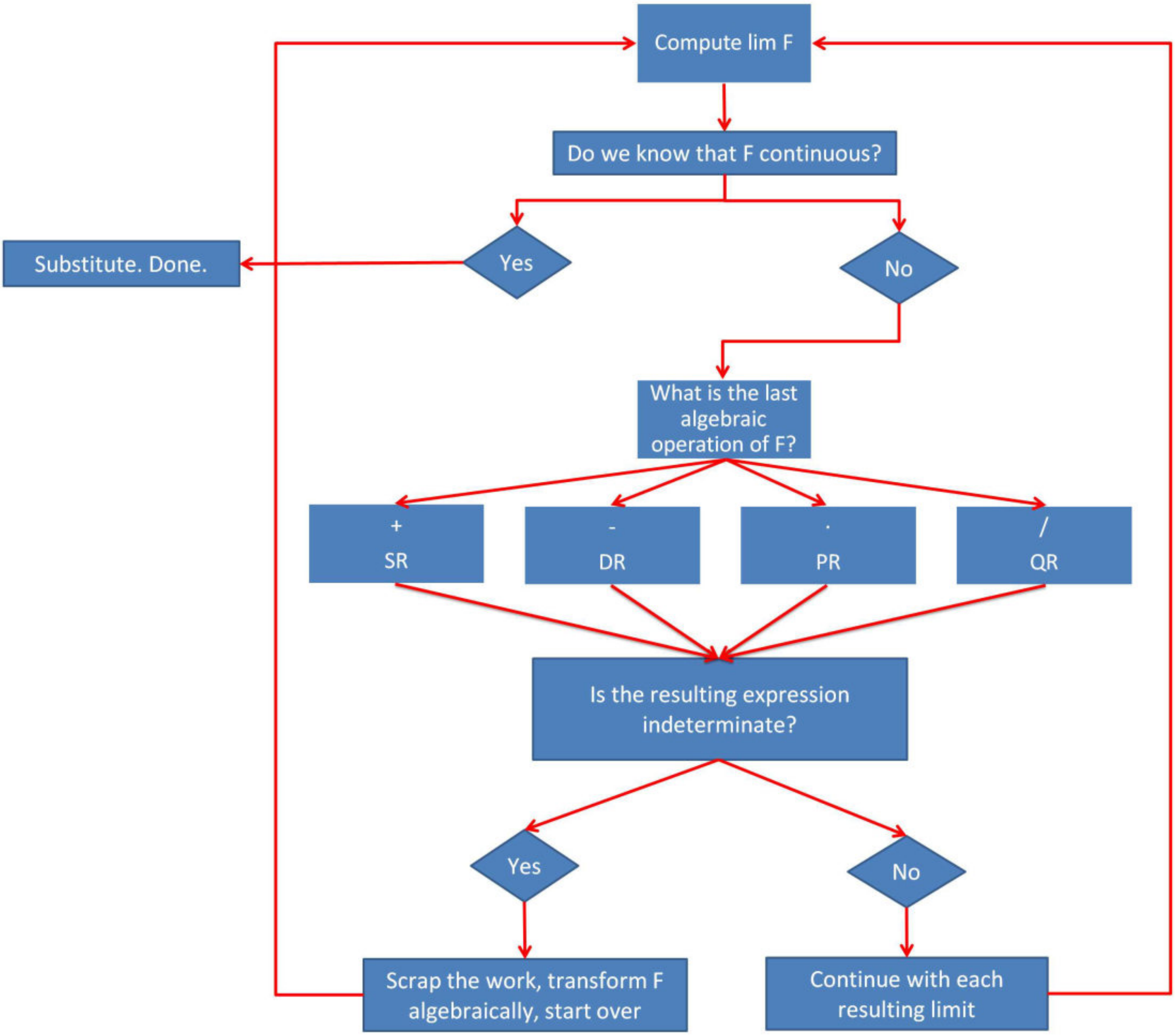
Therefore, its graph has four branches.

Exercise 2.4.33

Discuss the continuity of the following function:



This is a flowchart for limit computation:



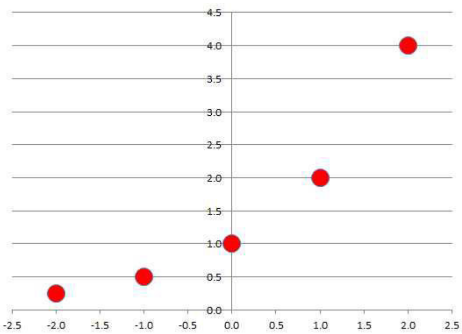
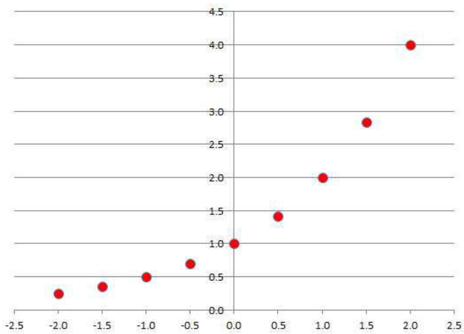
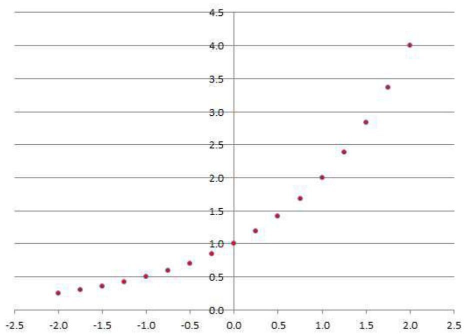
There is only one way out!

2.5. The exponential and trigonometric functions

Is the *exponential function* continuous?

Let's review how we construct it.

We have defined (in Volume 1, Chapters 1PC-1 and 1PC-4) the exponential function $f(x) = a^x,\ a > 0$ for all *integer* values of x as repeated multiplication and then also defined as repeated division in case of a negative x and then added more values by dividing these intervals in half as many times as necessary:

We used the *geometric mean* to define the value at this new point:

$$a^{\frac{x+y}{2}} = \sqrt{a^x \cdot a^y}\,.$$

We can also extend the definition to all *rational numbers* $x = \dfrac{p}{q}$ by means of the formula:

$$\boxed{a^{\frac{p}{q}} = \sqrt[q]{a^p} = \left(\sqrt[q]{a}\right)^p}$$

The gaps in the graph become invisible at once.

Now, what about $y = e^x$ for the rest of the *real* values of x? We use limits; for example,

$$e^{\pi} = \lim_{n\to\infty} e^{x_n},$$

for any sequence x_n of rational numbers that converges to π.

Of course, since there are infinitely many sequences converging to a number, we need to show that they all produce the same result!

We already have one limit from Chapter 1 that ensures a single outcome:

$$\boxed{\lim_{n\to\infty} e^{x_n} = 1}$$

for any sequence $x_n \to 0$. Now the rest of numbers:

Definition 2.5.1: exponential function for real argument

We define the values of the exponential function at $x = a$ as this limit:

$$\boxed{e^a = \lim_{n\to\infty} e^{x_n}}$$

for any sequence $x_n \to a$ of rational numbers.

We need to prove that the definition makes sense:

Theorem 2.5.2: Validity of Definition of Exponent

The definition of the exponential function does not depend on the choice of the sequence.

In other words, we have for any a:

$$\lim_{n\to\infty} e^{x_n} = e^a\,,$$

`FOR ANY` *sequence $x_n \to a$.*

Proof.

Let $h_n = x_n - a$. Then:

$$\begin{aligned}
\lim_{n\to\infty} e^{x_n} &= \lim_{n\to\infty} e^{a+h_n} \\
&= \lim_{n\to\infty} e^a e^{h_n} && \text{According to a rule of exponents.} \\
&= e^a \cdot \lim_{n\to\infty} e^{h_n} && \text{According to CMR.} \\
&= e^a \cdot 1 && \text{According to the above limit.} \\
&= e^a\,.
\end{aligned}$$

Let's rephrase the result in terms of the continuity of this function.

In terms of limits of functions, we first restate the limit from the last chapter:

$$\lim_{x\to 0} e^x = 1\,.$$

In other words:

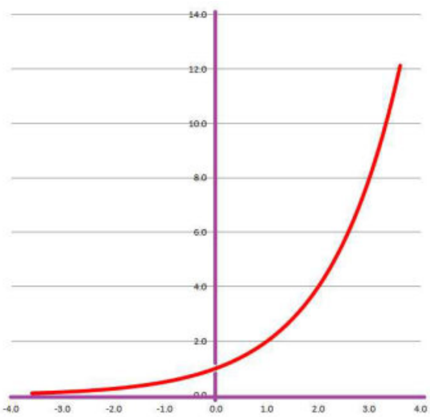
$$\lim_{x\to 0} e^x = e^0\,.$$

Therefore, we have proven the continuity of e^x at one single point, $x = 0$. We now derive the rest as follows:

Theorem 2.5.3: Continuity of Natural Exponential Function

The exponential function $y = e^x$ is continuous at every x.

The theorem confirms that the graphs of $y = e^x$ does indeed look like this, even if we zoom in on any point:



Exercise 2.5.4

Prove:

$$e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n\,.$$

Corollary 2.5.5: Continuity of Exponential Functions

Any exponential function (i.e., the exponential function $y = b^x$ of any base b) is continuous at every x.

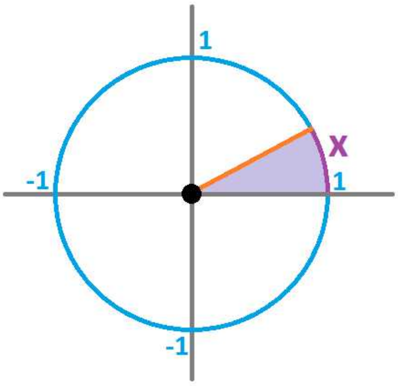
Exercise 2.5.6

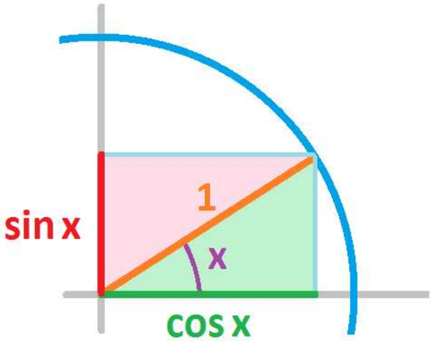
Prove the corollary.

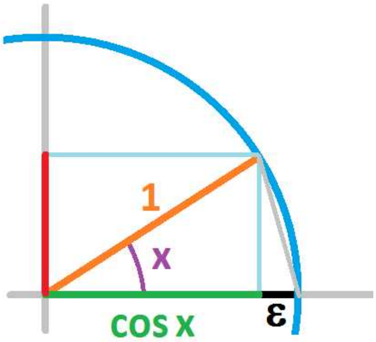
We follow the same path for the *trigonometric functions* and justify treating these functions as defined on $(-\infty, \infty)$.

Suppose a real number x is given. We construct a line segment of length 1 on the plane. Then:

- $\cos x$ is the horizontal coordinate of the end of the segment.
- $\sin x$ is the vertical coordinate of the end of the segment.

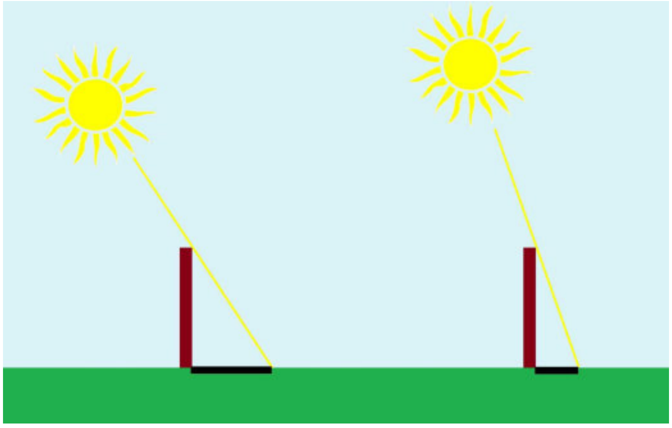




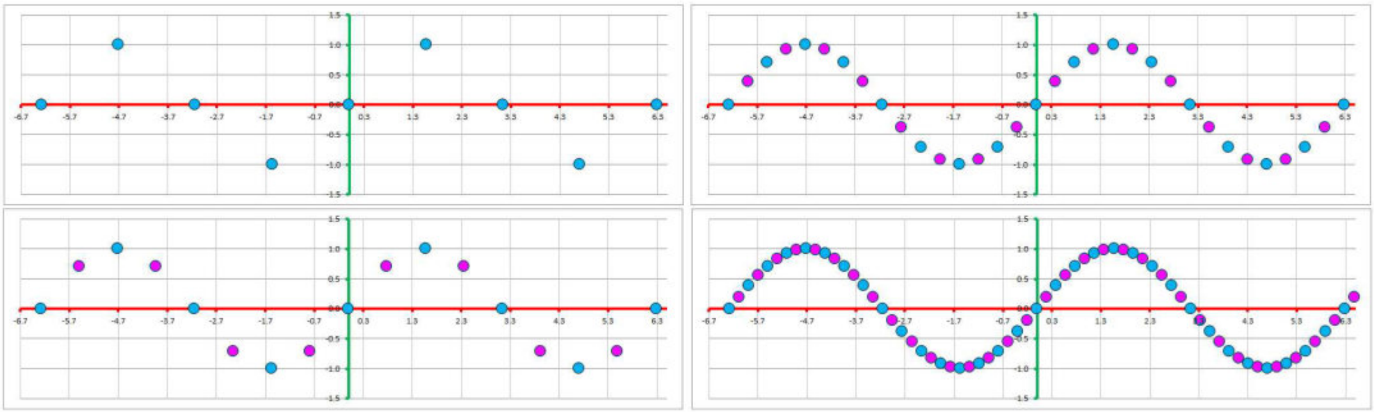


We think of $\cos x$ as the length of the shadow of the stick of length 1 under angle x in the ground when the sun is above it and $\sin x$ as the length of its shadow on the wall at sunset.

It is then plausible that – as the stick rotates – the length of the shadow changes continuously. Or the stick is still and it is the sun that is moving. Then the shadow gives us $\cos x$, where x is a multiple of time:



In the last chapter, we defined the sine and cosine algebraically by using the half-angle formula. The functions, therefore, are only defined on the fractions of π:



In other words, the domain is

$$X = \{r\pi : \ r \text{ rational }\}.$$

What about the rest?

What is $\sin 1$? We use limits; for example,

$$\sin 1 = \lim_{n\to\infty} \sin\big(\pi \cdot \big(1 + 1/n\big)\big)\,.$$

Of course, since there are infinitely many sequences converging to a number, we need to show that they all produce the same result. We already have the two limits from Chapter 1 that ensure a single outcome:

$$\boxed{\lim_{n\to\infty} \sin x_n = 0 \quad \text{and} \quad \lim_{n\to\infty} \cos x_n = 1}$$

for any sequence $x_n \to 0$. Now the rest of numbers:

Definition 2.5.7: Sine and Cosine

We define the values of the sine and the cosine at $x = a$ as these limits:

$$\boxed{\sin a = \lim_{n\to\infty} \sin x_n \quad \text{and} \quad \cos a = \lim_{n\to\infty} \cos x_n}$$

for any sequence $x_n \to a$ in X.

We need to prove that the definition makes sense:

Theorem 2.5.8: Validity of Definition of Sine and Cosine

The definition does not depend on the choice of the sequence.

In other words, we have:

$$\lim_{n\to\infty} \sin x_n = \sin a \quad \textit{and} \quad \lim_{n\to\infty} \cos x_n = \cos a\,,$$

for any sequence $x_n \to a$ and any a.

Proof.

Let $h_n = x_n - a$. Then:

$$\begin{aligned}
\lim_{n\to\infty} \sin(x_n) &= \lim_{n\to\infty} \sin(a + h_n) \\
&= \lim_{h\to 0} \left(\sin a \cdot \cos h + \cos a \cdot \sin h_n \right) && \text{A trig formula (Chapter 1PC-5).} \\
&= \lim_{n\to\infty} \left(\sin a \cdot \cos h_n \right) + \lim_{n\to\infty} \left(\cos a \cdot \sin h_n \right) && \text{According to SR.} \\
&= \sin a \cdot \lim_{n\to\infty} \cos h_n + \cos a \cdot \lim_{n\to\infty} \sin h_n && \text{According to CMR.} \\
&= \sin a \cdot 1 + \cos a \cdot 0 && \text{From the above limits.} \\
&= \sin a \,.
\end{aligned}$$

Exercise 2.5.9

Prove the theorem for the cosine.

Exercise 2.5.10

Sketch the graph of your elevation during a ride on the Ferris wheel.

Let's rephrase the result in terms of the continuity of these two functions.

In terms of limits of functions, we can restate the two limits from the last chapter:

$$\lim_{x\to 0} \sin x = 0 \quad \text{and} \quad \lim_{x\to 0} \cos x = 1 \,.$$

In other words:

$$\lim_{x\to 0} \sin x = \sin 0 \quad \text{and} \quad \lim_{x\to 0} \cos x = \cos 0 \,.$$

Therefore, we have already proven the continuity of the sine and the cosine at one single point, $x = 0$. We now derive the rest as follows:

Theorem 2.5.11: Continuity of Sine and Cosine

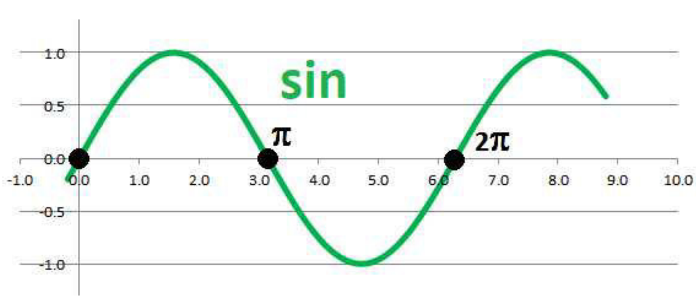
Both the cosine, $y = \cos x$, and the sine, $y = \sin x$, are continuous at every x.

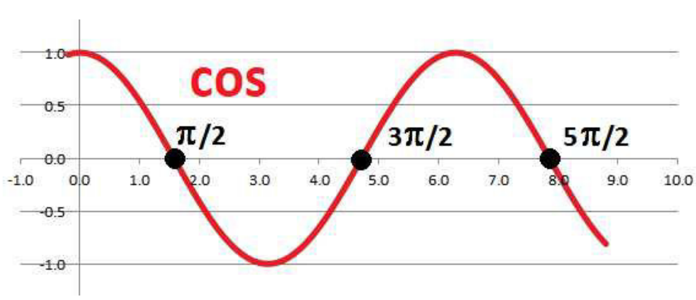
Proof.

We repeat the above proof:

$$\begin{aligned}
\lim_{h\to 0} \sin(a + h) &= \lim_{h\to 0} \left(\sin a \cdot \cos h + \cos a \cdot \sin h \right) && \text{A trig formula (Chapter 1PC-5).} \\
&= \lim_{h\to 0} \left(\sin a \cdot \cos h \right) + \lim_{h\to 0} \left(\cos a \cdot \sin h \right) && \text{According to SR.} \\
&= \sin a \cdot \lim_{h\to 0} \cos h + \cos a \cdot \lim_{h\to 0} \sin h && \text{According to CMR.} \\
&= \sin a \cdot 1 + \cos a \cdot 0 && \text{From the above limits.} \\
&= \sin a \,.
\end{aligned}$$

The theorem confirms that the graphs of $y = \sin x$ and $y = \cos x$ do indeed look like this, even if we zoom in on any point:





Below, we re-state the famous limits as facts about continuity.

Exercise 2.5.12

Show that the following functions are continuous at every point:

$$f(x) = \begin{cases} \dfrac{\sin x}{x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0 \end{cases} \qquad g(x) = \begin{cases} \dfrac{1-\cos x}{x} & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

We won't go into trigonometry beyond the tangent:

Corollary 2.5.13: Continuity of Tangent

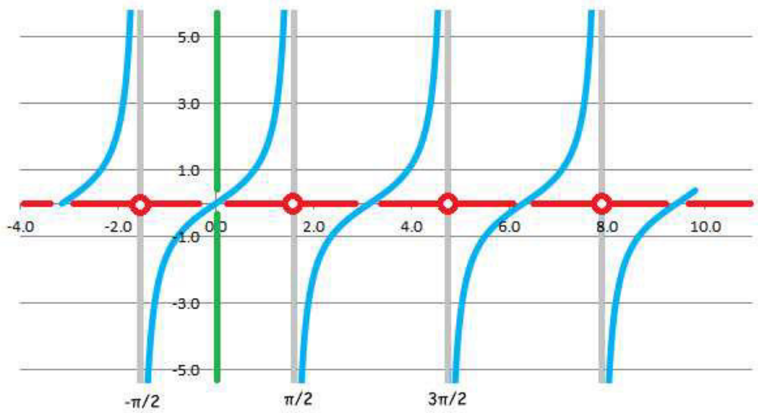
The tangent, $y = \tan x$, is continuous at every x where it is defined.

Proof.

Since
$$\tan x = \frac{\sin x}{\cos x}\,,$$
we apply the *Quotient Rule* to conclude that it is continuous at every point x with $\cos x \neq 0$.

Example 2.5.14: limit of tangent

Let's consider a point where the tangent is undefined, $x = \pi/2$:



We know that as $x \to \pi/2$, we have:

$$\sin x \to 1 \quad \text{and} \quad \cos x \to 0\,.$$

We should conclude that
$$\tan x \to \pm\infty\,.$$

However, which infinity? We take into account the *sign* of $\cos x$ on the two sides of $\pi/2$:

$$0 < \cos x \to 0 \text{ as } x \to \pi/2^- \quad \text{and} \quad 0 > \cos x \to 0 \text{ as } x \to \pi/2^+\,.$$

Therefore,
$$\tan x \to +\infty \text{ as } x \to \pi/2^- \quad \text{and} \quad \tan x \to -\infty \text{ as } x \to \pi/2^+\,.$$

Indeed, the graph reveals different behaviors on the two sides of $\pi/2$. The pattern repeats itself every π units on the x-axis.

We have added more functions to our collection of continuous functions!

In summary, all polynomials, all rational functions, the exponential functions, and the trigonometric func-

tions are continuous (on their domains). Therefore, their limits are evaluated by substitution:

$$\lim_{x\to 0} e^x = e^0,\ \lim_{x\to 0}\sin x = \sin 0,\ \text{etc.}$$

Combining these functions via the four algebraic operations will produce more continuous (on their domains) functions:

$$x^2+\sin x,\ \cos x\cdot e^x,\ \frac{x^3+3}{\sin x+2^x},\ \text{etc.}$$

However, what about these:

$$\sin(x^2+1),\ e^{\cos x},\ \text{etc. ?}$$

2.6. *Limits and continuity under compositions*

In the large collection of algebraic operations on functions, compositions are the most important.

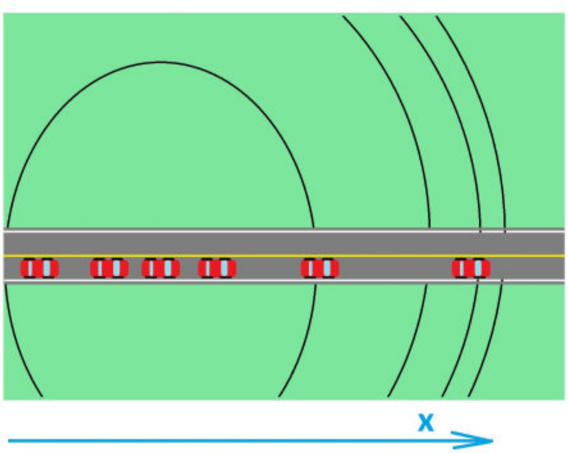
Example 2.6.1: driving through terrain

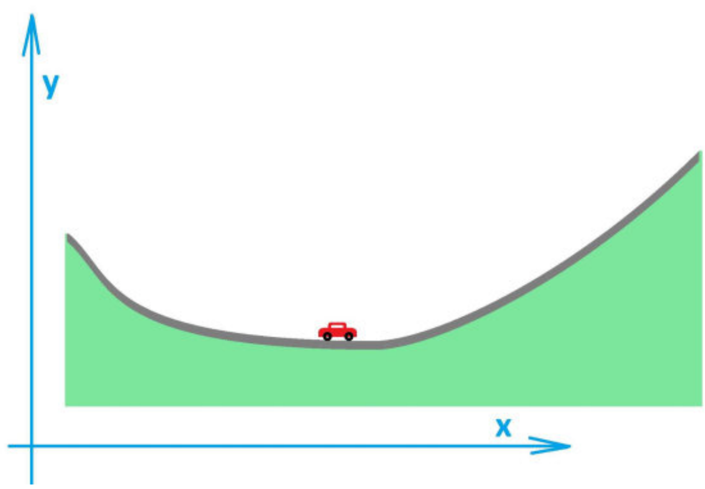
This is how compositions might emerge.

Suppose a car is driven through a mountain terrain and we know the following:

- We know where we are on the map at every moment of time.
- The map tells us the altitude for each location.

These two pieces of information are shown below as relation between the three quantities:



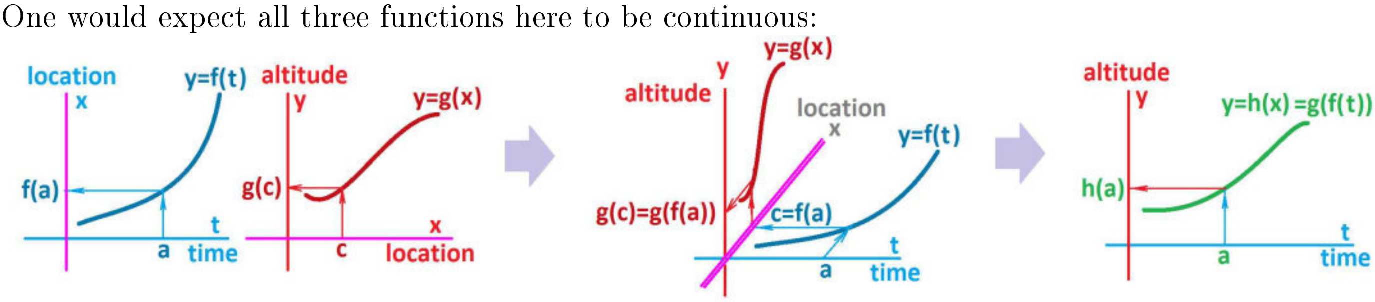


We set up two functions, for location and altitude, and their composition is what we are interested in:

- t is time – measured in hr
- $x=f(t)$ is the location of the car as a function of time – measured in mi
- $y=g(x)$ is the elevation of the road as a function of (horizontal) location – measured in ft
- $y=h(t)=g(f(t))$ is the altitude of the road as a function of time – measured in ft

The first function describes the motion, and the second function is literally the profile of the road.

One would expect all three functions here to be continuous:



In general, this is the setup:

Definition 2.6.2: composition of functions

Suppose we have two functions (with the codomain of the former matching the domain of the latter):

$$F : X \to Y \quad \text{and} \quad G : Y \to Z\,.$$

Then their *composition* is the function (from the domain of the former to the codomain of the latter)

$$H : X \to Z\,,$$

which is computed for every x in X according to the following two-step procedure:

$$x \ \to\ F(x) = y \ \to\ G(y) = z\,.$$

In other words, the new function is given by the *substitution formula*:

$$z = H(x) = G(F(x))\,.$$

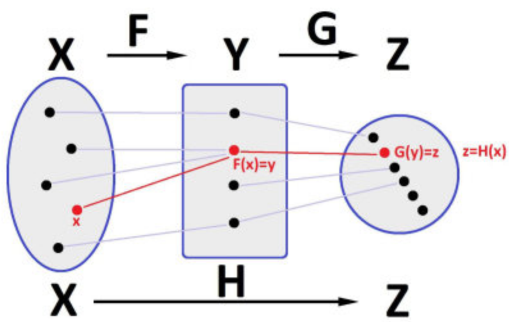
It is denoted as follows:

$$\boxed{G \circ F}$$

Warning!

We read compositions from right to left.

We just follow from X along the arrows of F to Y and then along the arrows of G to Z:



There may be more:

$$\boxed{X \xrightarrow{\ F\ } Y \xrightarrow{\ G\ } Z \xrightarrow{\ H\ } U \xrightarrow{\ Q\ } W \longrightarrow \ldots}$$

It's as if upon arriving at a location we are given directions to a new destination, on and on.

This is the "deconstruction" of the notation:

Composition of functions

names of the second and first functions

$$\begin{array}{ccc} & & \downarrow \quad \downarrow \\ (\ G \circ F\ )\ (x) & = & G(\quad F\ (x) \quad) \\ \uparrow & & \uparrow \quad \uparrow \quad \uparrow \end{array}$$

name of the new function — substitution

If we represent the two functions as *black boxes*, we can wire them together:

$$\begin{array}{|ccccc|}\hline \text{input} & & \text{function} & & \text{output}\\ x & \to & \boxed{\ F\ } & \to & y\\ \hline \end{array}$$

$$\downarrow$$

$$\begin{array}{|ccccc|}\hline \text{input} & & \text{function} & & \text{output}\\ y & \to & \boxed{\ G\ } & \to & z\\ \hline \end{array}$$

Here, we use the output of the former as the input of the latter, thus creating a new function:

$$\begin{array}{|ccccc|}\hline \text{input} & & & & \\ x & \to & \to & \to & \downarrow\\ \hline \end{array}$$

$$\begin{array}{|ccccc|}\hline & & & & \text{output}\\ \downarrow & \to & \to & \to & z\\ \hline \end{array}$$

To make it clear that Y is no longer a part of the picture, we can also visualize the composition as follows:

$$\begin{tikzcd} X \arrow[r, "F"] \arrow[dr, "H"] & Y \arrow[d, "G"] \\ & Z \end{tikzcd}$$

The meaning of the diagram is as follows: Whether we follow the F-then-G route or the direct H route, the results will be the same.

If we think of functions as *lists of instructions*, we just attach the list of the latter at the bottom of the list of the former. In other words, here is the list of $G \circ F$:

- Step 1: Do F.
- Step 2: Do G.

They are executed *consecutively*; you can't start with the second until you are done with the first.

The composition can also be written via the substitution notation:

$$\boxed{(g \circ f)(x) = g(u)\Big|_{u=f(x)}}$$

Here, u is the "intermediate" variable.

Example 2.6.3: composition of numerical functions

This function on the left is understood and evaluated via the diagram on the right:

$$f\,(y) = \frac{2y^2 - 3y + 7}{y^3 + 2y + 1}, \quad f\,(\square) = \frac{2\square^2 - 3\square + 7}{\square^3 + 2\square + 1}.$$

We can do the substitution $y = 3$ by inserting 3 in each of these windows:

$$f\left(\boxed{3}\right) = f(y)\Big|_{y=3} = \frac{2\boxed{3}^2 - 3\boxed{3} + 7}{\boxed{3}^3 + 2\boxed{3} + 1}.$$

Let's insert $\sin x$, or, better, $(\sin x)$. This is the result of the substitution $y = \sin x$:

$$f\boxed{(\sin x)} = f(y)\Big|_{y=\sin x} = \frac{2\boxed{(\sin x)}^2 - 3\boxed{(\sin x)} + 7}{\boxed{(\sin x)}^3 + 2\boxed{(\sin x)} + 1}.$$

Then, we have

$$f(\sin x) = (f \circ \sin)(x) = \frac{2(\sin x)^2 - 3(\sin x) + 7}{(\sin x)^3 + 2(\sin x) + 1}.$$

Exercise 2.6.4

Represent this function as a list of instructions:

$$f(x) = \left(\sqrt[3]{\sin x} + 2\right)^{1/2}.$$

Exercise 2.6.5

Find a formula for the following function:

$$\to \boxed{\text{square it}} \to \boxed{\text{take its reciprocal}} \to$$

Exercise 2.6.6

Suppose the graph of a function f is provided. How do you plot the graph of $y = -f(-x-3) - 1$?

Now, *limits and continuity.*

An application of *Product Rule* in a simple situation reveals a new shortcut:

$$\begin{aligned}\lim_{x\to a}\left[\left(f(x)\right)^2\right] &= \lim_{x\to a}\left[f(x)\cdot f(x)\right] \\ &= \lim_{x\to a} f(x) \cdot \lim_{x\to a} f(x) \quad \text{According to PR.} \\ &= \left(\lim_{x\to a} f(x)\right)^2. \quad \text{Under the assumption that limit exists.}\end{aligned}$$

It seems that we are saying that "the limit of the square of a function is the square of the limit of that function". Just like with the rest of the algebraic operations!

Furthermore, a repeated use of *Product Rule* produces a more general formula:

$$\lim_{x\to a}\left[f(x)^n\right] = \left(\lim_{x\to a} f(x)\right)^n,$$

for any natural number n. We conclude the following:

- ▶ The limit of the power is equal to the power of the limit.

Example 2.6.7: powers

The rule is useful:

$$\lim_{x\to 0}\left[(\sin x)^{20}\right] = \left(\lim_{x\to 0}[\sin x]\right)^{20} = (0)^{20} = 0.$$

Let's give this formula a new interpretation; it's a composition:

$$x \to \boxed{f} \to u \to \boxed{u^n} \to z$$

We have:

$$f(x)^n = g(f(x)) \text{ with } g(u) = u^n.$$

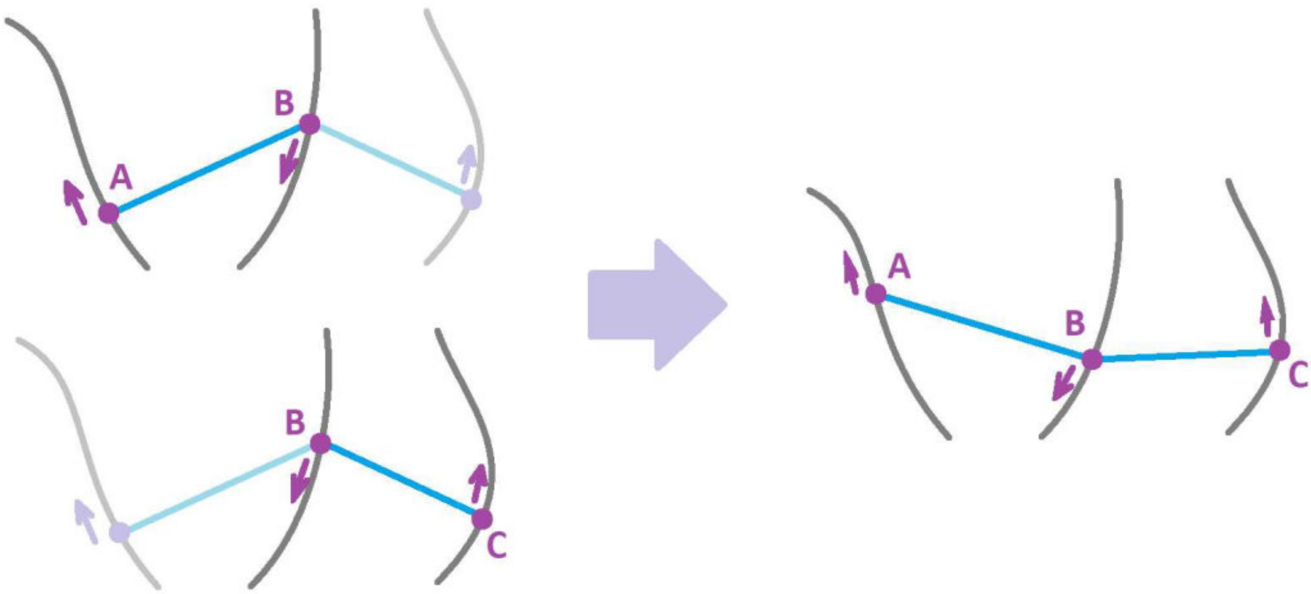
What is so special about this new function? It is continuous!

In brief,

▶ compositions of continuous functions are continuous.

Example 2.6.8: continuity of compositions, motion

The idea is illustrated as follows. Imagine we have three curved wires with a freely moving nut on each. The nuts are connected by two rods. Then, if the first nut is moved, it moves the second, and the second moves the third:



Either connection guarantees continuous motion, and so does their combination. Similarly, a driver controls the axle with the steering wheel in a multistage but continuous way.

Example 2.6.9: continuity of compositions, transformations

Imagine we carry out two transformations of a real line in a row:

Simply put, if we didn't tear the rope during the first stage, or during the second stage, then it's not torn!

We first consider the limit at a single point:

Theorem 2.6.10: Composition Rule for Limits

If the limit at a of function $f(x)$ exists, and is equal to L, then so does that of its composition with any function g continuous at L.

Furthermore, this limit is equal to $g(L)$:

$$\lim_{x\to a}(g\circ f)(x)=g(L)\,.$$

Proof.

Suppose we have a sequence,

$$x_n\to a\,.$$

Then, we also have another sequence,

$$b_n=f(x_n)\,.$$

The condition $f(x) \to L$ as $x \to a$ is restated as follows:

$$b_n \to L \text{ as } n \to \infty .$$

Therefore, the continuity of g implies,

$$g(b_n) \to g(L) \text{ as } n \to \infty .$$

In other words,

$$(g \circ f)(x_n) = g(f(x_n)) \to g(L) \text{ as } n \to \infty .$$

Since sequence $x_n \to a$ was chosen arbitrarily, this condition is restated as,

$$(g \circ f)(x) \to g(L) \text{ as } x \to a .$$

We can rewrite the result as a *substitution*:

$$\boxed{\lim_{x\to a}(g \circ f)(x) = g(L)\Bigg|_{L=\lim_{x\to a} f(x)}}$$

In other words, the continuous g can be moved out of the limit to be computed:

$$\boxed{\lim_{x\to a} g\big(f(x)\big) = g\left(\lim_{x\to a} f(x)\right)}$$

It's a simplification move.

Example 2.6.11: substitutions

Because these two functions are continuous, we can substitute twice:

$$\lim_{x\to 0} \sin\big(\cos x\big) = \sin\big(\lim_{x\to 0}\cos x\big) = \sin\big(\cos 0\big) = \sin\big(1\big) .$$

But this is the same as a single substitution:

$$\lim_{x\to 0} \sin\big(\cos x\big) = \lim_{x\to 0} \sin\circ\cos(x) = \sin\circ\cos(0) = \sin\big(1\big) .$$

The following is a crucial consequence of the theorem:

Corollary 2.6.12: Continuity under Compositions

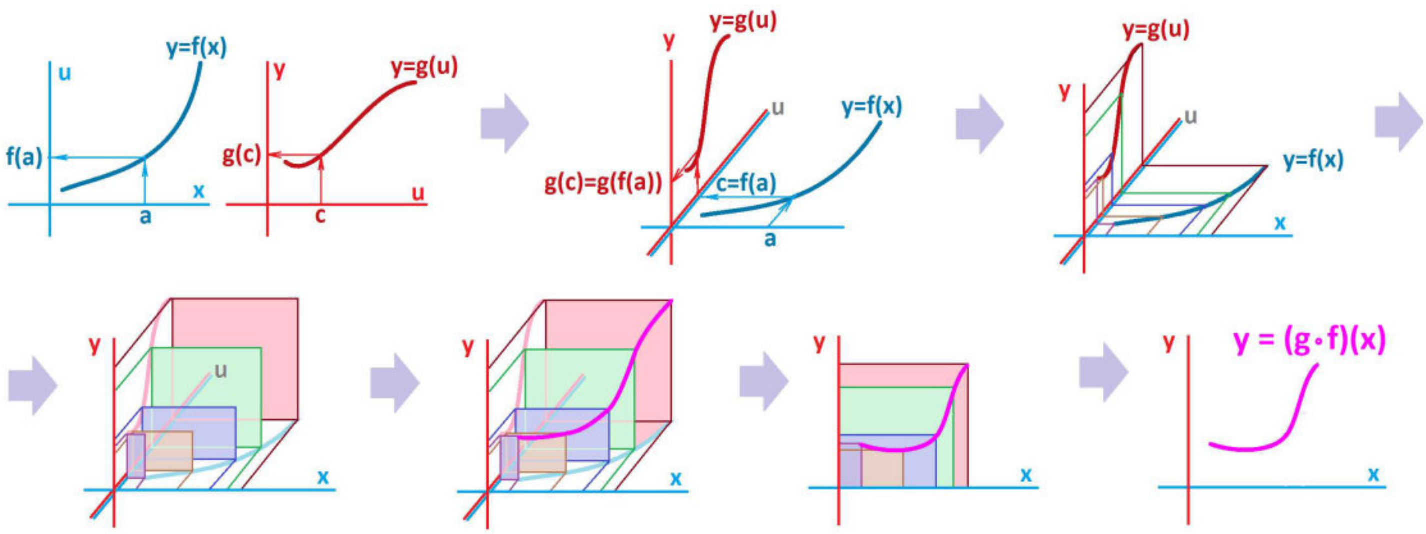
The composition $g \circ f$ of a function f continuous at $x = a$ and a function g continuous at $y = f(a)$ is continuous at $x = a$.

Proof.

We apply the theorem and then the continuity of f:

$$\lim_{x\to a}(g \circ f)(x) = g\left(\lim_{x\to a} f(x)\right) = g\left(f(a)\right) = (g \circ f)(a) .$$

The geometry of the composition of continuous functions is illustrated below:



Now algebra. This is how two continuous functions interact. First, we have the continuity of either described separately:

- 1. A small deviation of x from a produces a small deviation of $u = f(x)$ from $f(a)$.
- 2. A small deviation of u from c produces a small deviation of $y = g(u)$ from $g(c)$.

If we set $c = f(a)$, we have:

- 3. A small deviation of $u = f(x)$ from $c = f(a)$ produces a small deviation of $y = g(u) = g(f(x))$ from $g(c) = g(f(a))$.

That's continuity of $h = g \circ f$ at $x = a$! The diagram illustrates this observation:

$$\begin{array}{ccc} \boxed{x \to a} & \xrightarrow{f} & \boxed{f(x) \to f(a)} \\ & \searrow^{g\circ f} & \downarrow g \\ & & \boxed{g(f(x)) \to g(f(a))} \end{array}$$

Example 2.6.13: $\sin$ and $\cos$

From trigonometry, we know that

$$\cos(x) = \sin(\pi/2 - x)\,.$$

Therefore, the continuity of sin implies the continuity of cos, and vice versa.

Example 2.6.14: algebra of compositions

Consider these two functions and their composition:

$$\begin{array}{ll} y = g(u) & = u^2 + 2u - 1\,, \\ u = f(x) & = 2x^{-3}\,. \end{array}$$

What is the limit of $h = g \circ f$ at 1?

First, we note that g is continuous at every point as a polynomial. Therefore, by the theorem we have:

$$\lim_{x\to 1} (g \circ f)(x) = g\left(\lim_{x\to 1} f(x)\right)\,,$$

if the limit on the right exists. It does, because f is a rational function defined at $x = 1$:

$$\lim_{x\to 1} f(x) = \lim_{x\to 1} 2x^{-3} = 2 \cdot 1^{-3} = 2\,.$$

The limit becomes a number, $u = 2$, and this number is substituted into g:

$$\lim_{x\to 1} h(x) = g\left(\lim_{x\to 1} f(x)\right) = g(2) = 2^2 + 2 \cdot 2 - 1 = 7\,.$$

The answer is verified by a direct computation of h:

$$h(x) = (g \circ f)(x) = g(f(x)) = u^2 + 2u - 1\Big|_{u=2x^{-3}} = \left(2x^{-3}\right)^2 + 2\left(2x^{-3}\right) - 1 = 4x^{-6} + 4x^{-3} - 1\,.$$

Since the function is continuous (rational and defined at 1), we have by substitution:

$$\lim_{x\to 1} h(x) = 4x^{-6} + 4x^{-3} - 1\Big|_{x=1} = 4 \cdot 1^{-6} + 4 \cdot 1^{-3} - 1 = 7\,.$$

Example 2.6.15: using the rule

Compute:

$$\lim_{x\to 0} \frac{1}{\left(x^2 + x - 1\right)^3}\,.$$

We proceed by a gradual decomposition of

$$r(x) = \frac{1}{\left(x^2 + x - 1\right)^3}\,.$$

The last operation of f is division. Therefore,

$$r(x) = g(f(x)), \text{ where } u = f(x) = \left(x^2 + x - 1\right)^3 \text{ and } g(u) = \frac{1}{u}\,.$$

Function g is rational; but is it continuous? Its denominator isn't zero at the point we are interested in:

$$f(0) = \left(x^2 + x - 1\right)^3\Big|_{x=0} = \left(0^2 + 0 - 1\right)^3 = -1 \neq 0\,.$$

Then the *Composition Rule for Limits* applies and our limit becomes:

$$\lim_{x\to 0} \frac{1}{\left(x^2 + x - 1\right)^3} = \lim_{x\to 0} r(x) = \lim_{x\to 0} g(f(x)) = g(\lim_{x\to 0} f(x)) = \frac{1}{\lim_{x\to 0}\left[\left(x^2 + x - 1\right)^3\right]}\,,$$

provided the new limit exists. Notice that the limit to be computed has been simplified!

Let's compute it. We start over and continue with a decomposition of

$$p(x) = \left(x^2 + x - 1\right)^3\,.$$

The last operation of f is the power. Therefore,

$$p(x) = g(f(x)), \text{ where } u = f(x) = x^2 + x - 1 \text{ and } g(u) = u^3\,.$$

Function g is a polynomial and, therefore, continuous. Then the *Composition Rule for Limits* applies and the limit becomes:

$$\lim_{x\to 0} \left(x^2 + x - 1\right)^3 = \lim_{x\to 0} p(x) = \lim_{x\to 0} g(f(x)) = g(\lim_{x\to 0} f(x)) = \left[\lim_{x\to 0} \left(x^2 + x - 1\right)\right]^3\,,$$

provided the new limit exists. Notice that, again, the limit to be computed has been simplified!

Let's compute it. We realize that the function $x^2 + x - 1$ is a polynomial and, therefore, its limit is computed by substitution:

$$\lim_{x\to 0} \left(x^2 + x - 1\right) = x^2 + x - 1\Big|_{x=0} = 0^2 + 0 - 1 = -1\,.$$

What remains is to combine the three formulas above:

$$\begin{aligned}
\lim_{x\to 0} \frac{1}{\left(x^2 + x - 1\right)^3} &= \frac{1}{\left[\lim_{x\to 0}\left(x^2 + x - 1\right)^3\right]} \\
&= \frac{1}{\left[\lim_{x\to 0}\left(x^2 + x - 1\right)\right]^3} \\
&= \frac{1}{[-1]^3} \\
&= -1\,.
\end{aligned}$$

When there is no continuity to be used, we may have to apply algebra (such as factoring), or trigonometry, etc. in order to find another decomposition of the function.

Our short list of continuous functions allows us to compute many limits; we just have to justify, every time, our conclusion by making a reference to this fact.

Example 2.6.16: computations

First,

$$\lim_{n\to\infty} \ln\left(1+\frac{1}{n}\right) = \ln\left(\lim_{n\to\infty}\left[1+\frac{1}{n}\right]\right) = \ln 1 = 0\,,$$

because $\ln x$ is continuous at $x=1$. Second,

$$\lim_{n\to\infty} \left(1+\frac{1}{n}\right)^2 = \left(\lim_{n\to\infty}\left[1+\frac{1}{n}\right]\right)^2 = 1^2 = 1\,,$$

because x^2 is continuous at $x=1$. Third,

$$\lim_{n\to\infty} \frac{1}{1+\frac{1}{n}} = \frac{1}{\lim_{n\to\infty}\left[1+\frac{1}{n}\right]} = \frac{1}{1} = 1\,,$$

because $\dfrac{1}{x}$ is continuous at $x=1$.

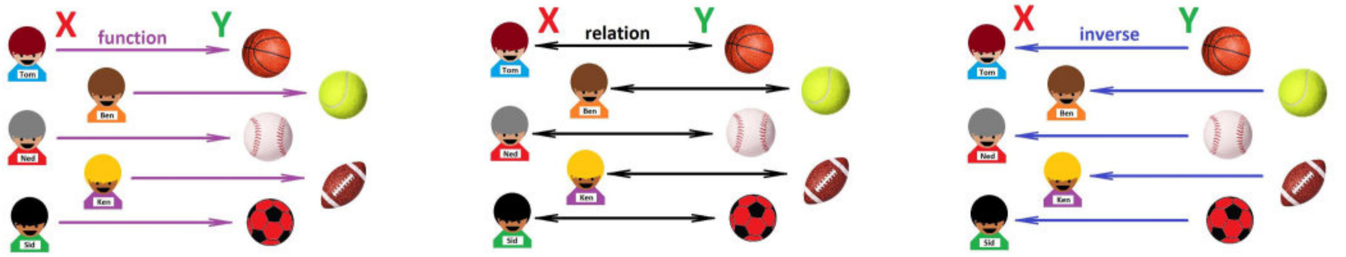
2.7. Continuity of the inverse

So far, the only large class of functions we know to be continuous is the rational functions: All algebraic operations, including compositions, on rational functions produce more rational functions. The continuity of such a simple function as the square root $\sqrt{x}$ remains unproven. But to see how we should approach the problem we just need to remember where the function comes from. It's the *inverse* of the squaring function y^2!

Let's review.

Example 2.7.1: relations

The inverse of a function is based on the same relation but with domain and codomain (or input and output) reversed:



Example 2.7.2: inverses

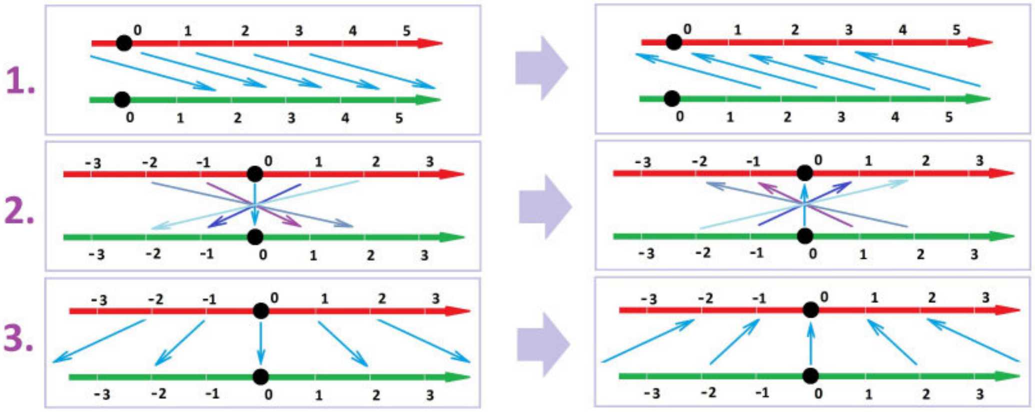
We also like to think of two functions as inverses when one *undoes the effect of the other*. For numerical functions, these are a few examples. A function is undone by its inverse, and vice versa:

	function	relation	inverse function	
multiplication by 3	$1 \to 3$	$1 \leftrightarrow 3$	$1 \leftarrow 3$	division by 3
squaring	$2 \to 4$	$2 \leftrightarrow 4$	$2 \leftarrow 4$	square root (for $x, y \geq 0$)
exponential function base 3	$2 \to 9$	$2 \leftrightarrow 9$	$2 \leftarrow 9$	logarithm base 3

But why is the inverse continuous?

Example 2.7.3: inverse of transformation

The basic transformations of the line provide more examples:



In other words, we have:

- The inverse of the shift s units to the right is the shift of s units to the left.
- The inverse of the flip is another flip.
- The inverse of the stretch by $k \neq 0$ is the shrink by k (i.e., stretch by $1/k$).

When executed consecutively, the effect is nil. Algebraically:

	f	vs. f^{-1}
shift	$y = x + s$	$x = y - s$
flip	$y = -x$	$x = -y$
stretch	$y = x \cdot k$	$x = y/k$

Within this class, it seems that its inverse should be continuous if the transformation is.

Let's recall the precise definition:

Definition 2.7.4: inverse of function

Functions $F : X \to Y$ and $G : Y \to X$ are called *inverse* of each other if they come from the same relation; i.e., for all x and y, we have the following:

$$F(x) = y \quad \texttt{IF AND ONLY IF} \quad G(y) = x\,.$$

The condition is illustrated with this diagram:

$$x \xrightarrow{\quad F \quad} y \quad \Longleftrightarrow \quad y \xrightarrow{\quad G \quad} x$$

The notation for G is as follows:

Inverse function

$$F^{-1}$$

It reads "F inverse".

Then we have a literal reversal of arrows:

$$\begin{array}{ccc} x & \xrightarrow{F} & y \\ x & \xleftarrow{F^{-1}} & y \end{array}$$

Exercise 2.7.5

Find the inverse of the function $f(x) = 3x^2 + 1$. Choose appropriate domains for these two functions.

Exercise 2.7.6

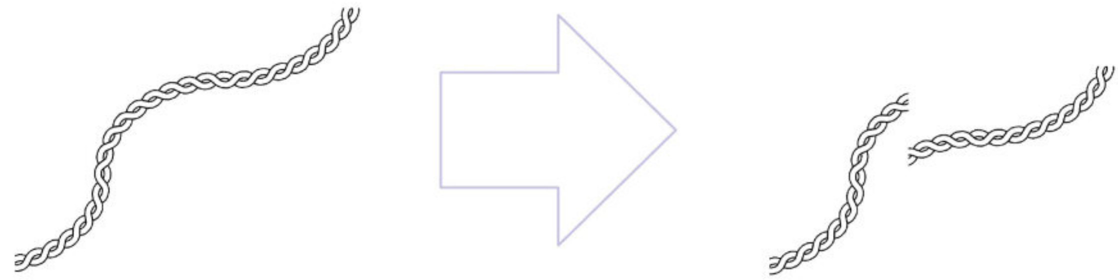
(a) Algebraically, show that the function $f(x) = x^2$ is not one-to-one. (b) Graphically, show that the function $g(x) = 2^{x+1}$ is one-to-one. (c) Find the inverse of g.

Now continuity.

Example 2.7.7: tearing

First, the inverses of the transformations above are continuous. What about others? How do we know that undoing a transformation doesn't tear the rope?

Well, what is the inverse of tearing? It's gluing!



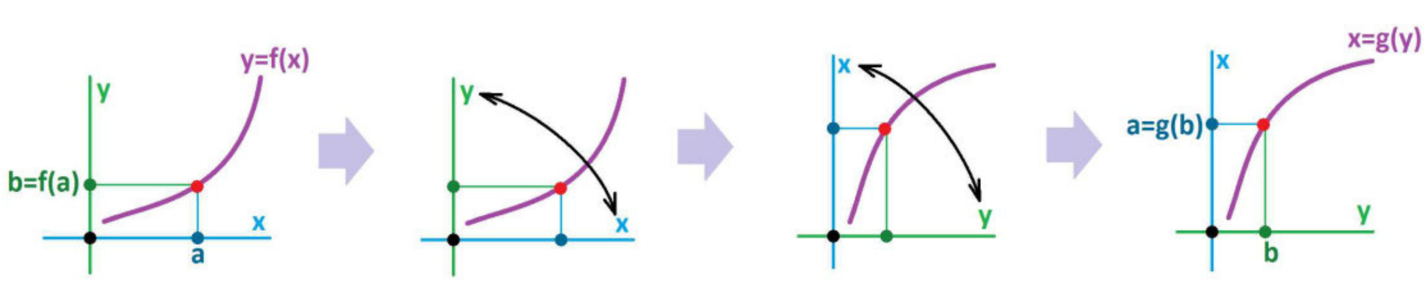
If we make sure that our original transformation doesn't include gluing of the pieces of the rope together, the inverse will be continuous too. But this simply means that the function has to be one-to-one.

The inverses undo each other *under composition*; functions $y = f(x)$ and $x = g(y)$ are called inverse of each other when for all x in the domain of f and for all y in the domain of g, we have:

$$\boxed{g(f(x)) = x \quad \texttt{AND} \quad f(g(y)) = y}$$

What if one of these functions, say f, is continuous? Does it make the other, g, continuous too? Considering the fact that the right-hand sides of these equations (x and y) are continuous functions, we expect the answer to be Yes.

Every function $y = f(x)$ which is *one-to-one* (i.e., there is only one x for each y) has the inverse $x = g(y)$ which is also one-to-one (i.e., there is only one y for each x). If we take the graph of the former and flip this piece of paper so that the x-axis and the y-axis are interchanged, we get the graph of the latter (and vice versa):



The shapes of the graphs are exactly the same! If one has no breaks, then neither does the other. It is then conceivable that the continuity of one implies the continuity of the other.

To link the continuity of the two functions, we have to be careful about the location of the sequence. We know that according to the definition, we have the following implication:

$$\boxed{x_n \to a \implies f(x_n) \to f(a)}$$

That's the assumption. Now, in reverse:

$$\boxed{f^{-1}(y_n) \to f^{-1}(b) \impliedby y_n \to b}$$

That's what we need to prove. We accept the following important conclusion without proof:

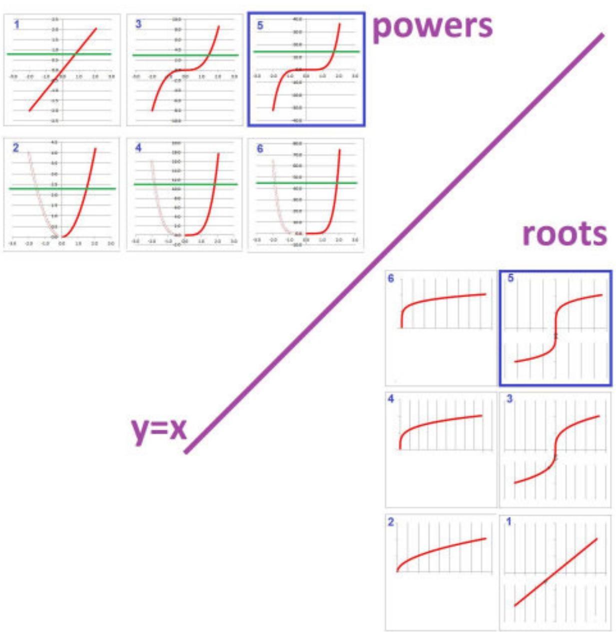
Theorem 2.7.8: Continuity of Inverse

The inverse of a function $y = f(x)$ continuous at $x = a$, if it exists, is a function $x = g(y)$ continuous at $b = f(a)$.

Thus, as the rational functions are continuous (within their domains), so are their inverses. In particular, the square root function is continuous and we have:

$$\lim_{x\to a} \sqrt{f(x)} = \sqrt{\lim_{x\to a} f(x)}\,.$$

What about the rest of power functions? Recall that since the odd powers are one-to-one but the even powers aren't one-to-one, we had to remove half of the domains of the latter:



The look of the new graph suggest the following:

Theorem 2.7.9: Continuity of Roots

- *The odd roots, $y = \sqrt[3]{x}$, $y = \sqrt[5]{x}$, $y = \sqrt[7]{x}$, ... are continuous at every real x.*
- *The even roots, $y = \sqrt[2]{x}$, $y = \sqrt[4]{x}$, $y = \sqrt[6]{x}$, ... are continuous at every $x \geq 0$.*

Another important pair of inverses is the exponent and the *logarithm*. Since the former is continuous, then so is the latter. This makes possible the following definitions:

As we know (as seen in Volume 1, Chapter 1PC-4), this function is increasing for all rational x. Then, from the *Comparison Theorem*, we conclude that it is also increasing on $(-\infty, \infty)$. Therefore, it is one-to-one.

Definition 2.7.10: natural logarithm

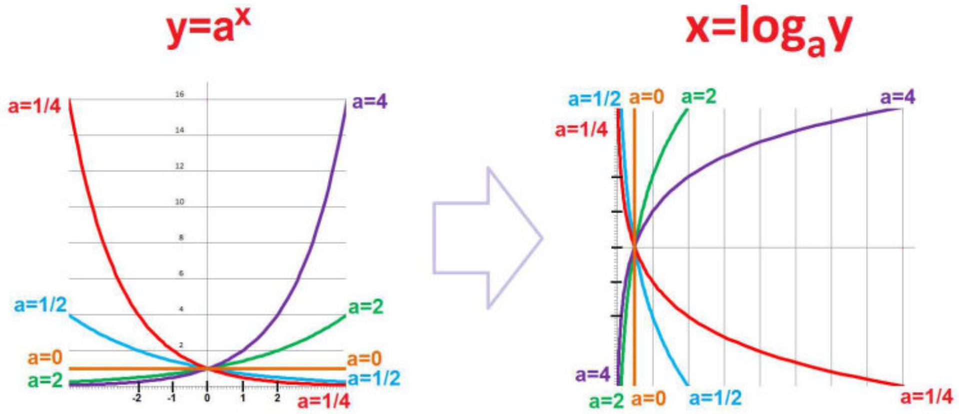
The *natural logarithm function*, or the logarithm function base e, is defined as the inverse of the natural exponential function.

Definition 2.7.11: exponential function

The *exponential function base* $a > 0$ is defined by its values for each real x:

$$a^x = e^{x \ln a}.$$

The graphs are just flipped:



We already know the following:

Theorem 2.7.12: Continuity of Exponential Function

The exponential function of any base $y = a^x$ is continuous.

Proof.

It follows from the continuity of $y = e^x$ and the *Composition Rule for Limits*.

We then derive the following about its inverse:

Theorem 2.7.13: Continuity of Logarithm

The logarithm of any base $y = \log_a x$ is continuous (on its domain $x > 0$).

The analysis of the trigonometric functions is similar to that for powers: we need to reduce the domains, as follows:

Definition 2.7.14: arcsine and arccosine

The *arcsine* is defined to be the inverse of the sine function restricted to $[-\pi/2, \pi/2]$, denoted by either of these:

$$\boxed{\arcsin y = \sin^{-1} y}$$

The *arccosine* is defined to be the inverse of the cosine function on $[0, \pi]$, denoted

by either of these:

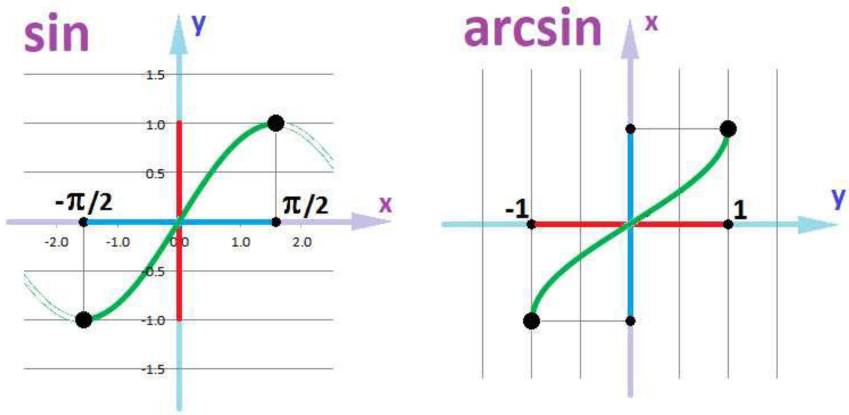
$$\boxed{\arccos y = \cos^{-1} y}$$

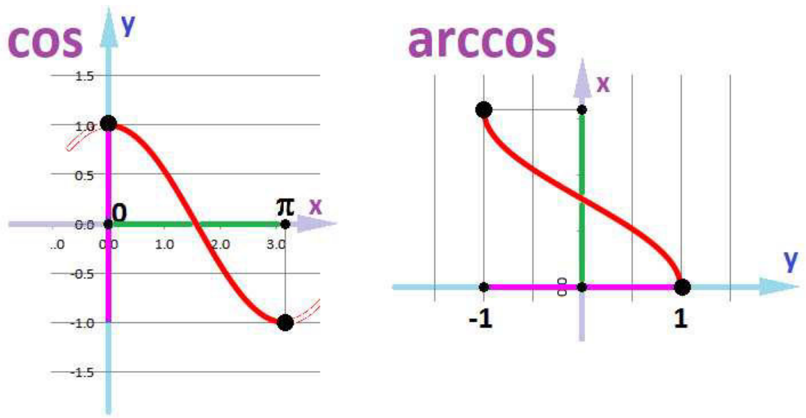
Thus, we have two pairs of inverse functions:

$y = \sin x$	$x = \sin^{-1} y$
domain: $[-\pi/2, \pi/2]$	domain: $[-1, 1]$
range: $[-1, 1]$	range: $[-\pi/2, \pi/2]$

$y = \cos x$	$x = \cos^{-1} y$
domain: $[0, \pi]$	domain: $[-1, 1]$
range: $[-1, 1]$	range: $[0, \pi]$

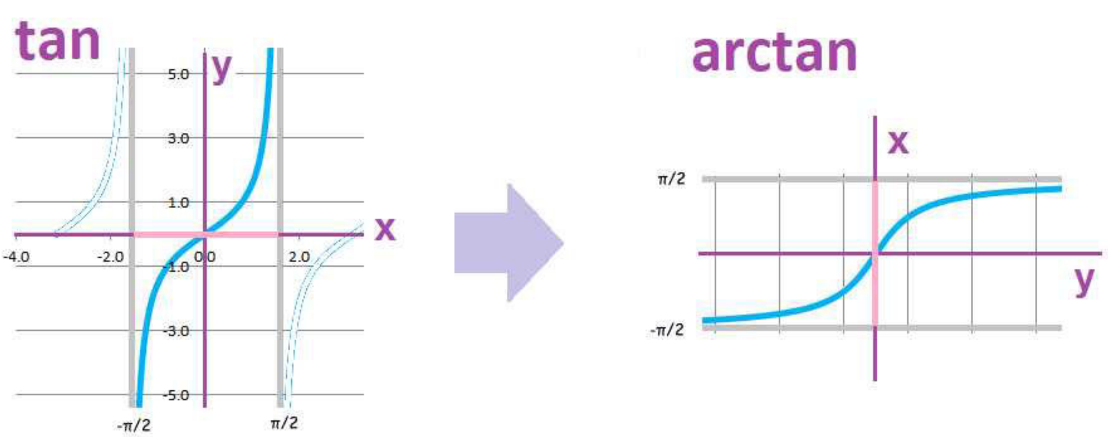
The graphs are, of course, the same with just x and y interchanged:





Exercise 2.7.15

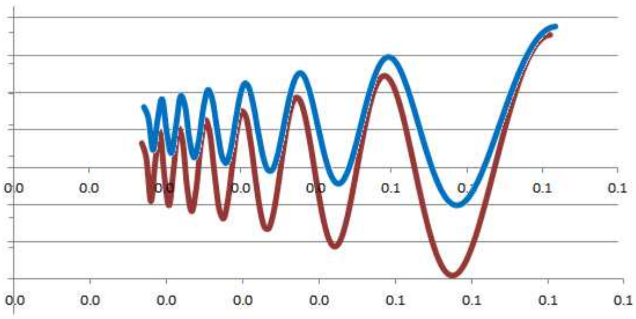
Conduct the same analysis for the tangent.



2.8. Comparison of limits

To show continuity of functions beyond the rational functions, we will employ some indirect and direct methods.

When two function are comparable in value, the limit of one might tell us something about that for the other:



The picture is an illustration of the following fact:

Theorem 2.8.1: Comparison Test for Limits of Functions

Non-strict inequalities between functions are preserved under limits.

In other words, if

$$f(x) \le g(x)$$

for all n greater than some N, then

$$\lim_{x\to a} f(x) \le \lim_{x\to a} g(x),$$

provided the limits exist.

Exercise 2.8.2

Prove the theorem.

Warning!

The theorem does not claim convergence.

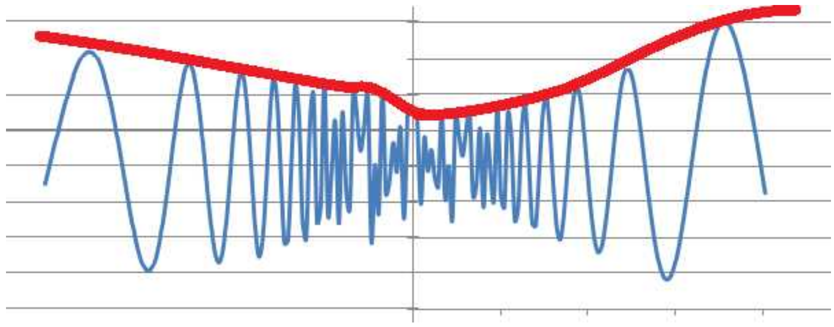
This is the summary of the theorem:

$$\begin{array}{cccc} & f(x) & \ge & g(x) \\ & \downarrow & & \downarrow \\ & L & & M \\ \Longrightarrow & & \ge & \end{array}$$

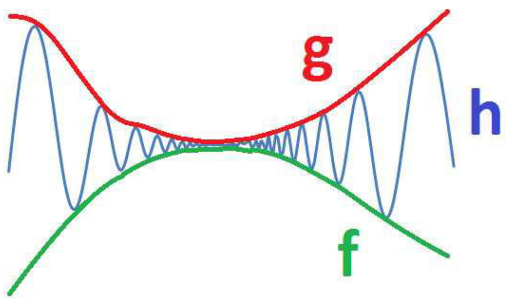
Exercise 2.8.3

Show that replacing the non-strict inequality $f(x) \le g(x)$ with a strict one $f(x) < g(x)$ won't produce a strict inequality in the conclusion of the theorem.

From this (single) inequality, we can't conclude anything about the existence of the limit:



So, even when two function are comparable in value, the convergence or divergence of one won't tell us anything about that of the other. Having *two* inequalities, on both sides, may work better:



Such a double inequality is called a *squeeze*, just as in the case of sequences. If we can squeeze the function under investigation between two familiar functions, we might be able to say something about its limit. Some further requirements will be necessary:

Theorem 2.8.4: Squeeze Theorem for Functions

If the values of a function lie between those of two functions with equal limit, then its limit also exists and is equal to that number.

In other words, if we have:

$$f(x) \le h(x) \le g(x)$$

for all x within some open interval that contains $x = a$, and if we have:

$$\lim_{x\to a} f(x) = \lim_{x\to a} g(x) = L\,,$$

then the following limit exists and is equal to the same number:

$$\lim_{x\to a} h(x) = L\,.$$

Proof.

For any sequence $x_n \to a$, we have:

$$f(x_n) \le h(x_n) \le g(x_n)\,.$$

We also have:

$$\lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} g(x_n) = L\,.$$

Therefore, by the *Squeeze Theorem for Sequences* we have:

$$\lim_{n\to\infty} h(x_n) = L\,.$$

Warning!

The theorem does claim convergence.

Example 2.8.5: find a squeeze

Let's find the limit,

$$\lim_{x\to 0} x \sin\left(\frac{1}{x}\right) \quad =?$$

It cannot be computed by the *Product Rule* because

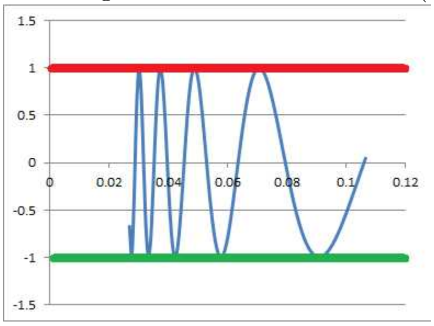
$$\lim_{x\to 0} \sin\left(\frac{1}{x}\right) \quad \text{does not exist.}$$

Let's try to find a squeeze instead.

This is what we know from trigonometry:

$$-1 \le \sin\left(\frac{1}{x}\right) \le 1\,.$$

Let's note that this squeeze proves nothing about the limit of $\sin(1/x)$:

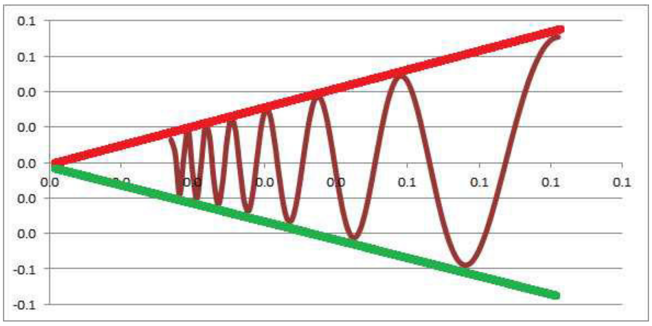


It's just two bounds!

Let's consider this squeeze produced from the last one:

$$-|x| \le x \sin\left(\frac{1}{x}\right) \le |x|\,.$$

It looks more useful:



Now, since $\lim\limits_{x\to 0}(-x) = \lim\limits_{x\to 0}(x) = 0$, by the *Squeeze Theorem*, we have:

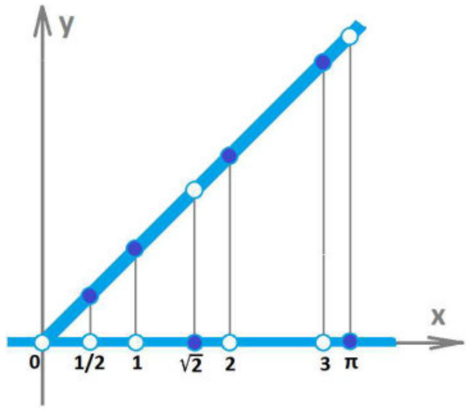
$$\lim_{x\to 0} x \sin\left(\frac{1}{x}\right) = 0\,.$$

Example 2.8.6: Dirichlet

We can use the same logic to prove that the Dirichlet function multiplied by x is continuous at exactly one point!

$$xI_Q(x) = \begin{cases} x & \text{if } x \text{ is a rational number,} \\ 0 & \text{if } x \text{ is an irrational number.} \end{cases}$$

Again, to plot its graph, we can only draw two lines and then point out *some* of the missing points:



The example shows that the absolute value is a good tool for constructing a squeeze. We apply this idea to all functions:

Corollary 2.8.7: Limit of Absolute Value

The limit of a function is zero if and only if that of its absolute value is.

In other words, we have:

$$\lim_{x\to a} f(x) = 0 \iff \lim_{x\to a} |f(x)| = 0\,.$$

Exercise 2.8.8

Prove the corollary.

Example 2.8.9: famous limit

The first of the two famous limits:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1 \quad \text{and} \quad \lim_{x\to 0} \frac{1-\cos x}{x} = 0\,,$$

follows by the *Squeeze Theorem* from this trigonometric fact:

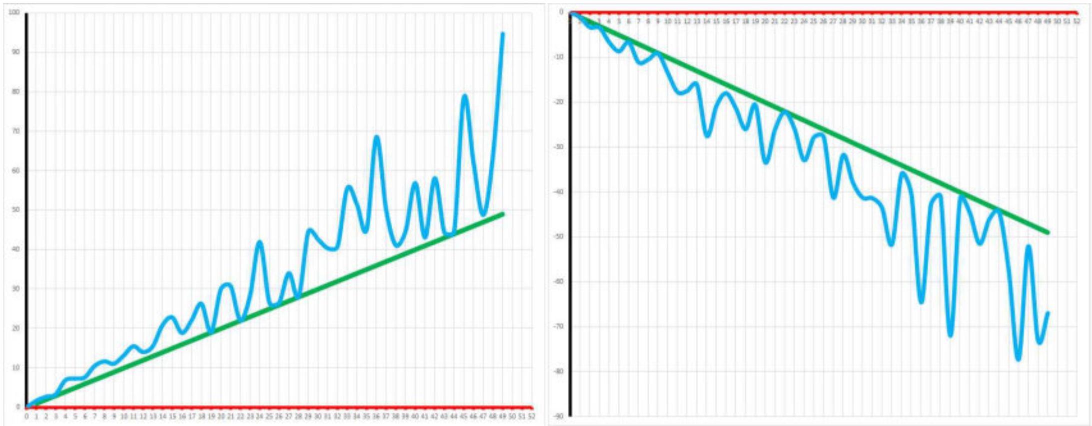
$$\cos x \le \frac{\sin x}{x} \le 1\,.$$

Exercise 2.8.10

Prove the other one in the same manner.

We now turn to *divergence*.

To make conclusions about divergence to *infinity*, we only need to control the function from *one*, but the right, side:



The smaller function will push the larger to the positive infinity and the larger function will push the smaller to the negative infinity. Below is the analog of the Squeeze Theorem for infinite limits:

Theorem 2.8.11: Push Out Theorem for Limits of Functions

- *If the smaller function approaches positive infinity, then so does the larger one.*
- *If the larger function approaches negative infinity, then so does the smaller one.*

In other words, if $f(x) \le g(x)$ for all x within some open interval that contains a, then we have:

$$\begin{array}{lcl} \lim\limits_{x\to a} f(x) = +\infty & \Longrightarrow & \lim\limits_{x\to a} g(x) = +\infty \\ \lim\limits_{x\to a} f(x) = -\infty & \Longleftarrow & \lim\limits_{x\to a} g(x) = -\infty \end{array}$$

Exercise 2.8.12

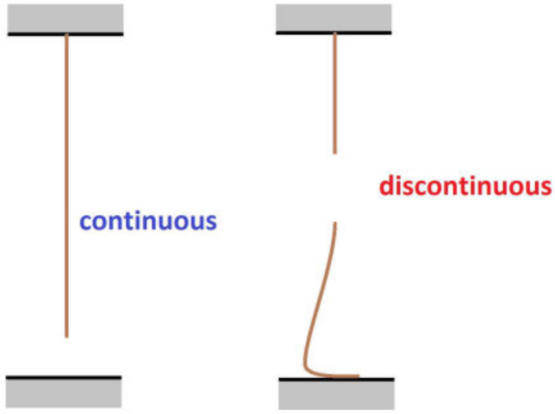
Suggest an example of how this theorem applies.

This is the summary of the theorem:

$$\begin{array}{cccc|cccc} & f(x) & \le & g(x) & & f(x) & \le & g(x) \\ & \downarrow & & & & & & \downarrow \\ & +\infty & & & & & & -\infty \\ \Longrightarrow & & & \downarrow & \Longrightarrow & \downarrow & & \\ & & & +\infty & & -\infty & & \end{array}$$

2.9. Global properties of continuous functions

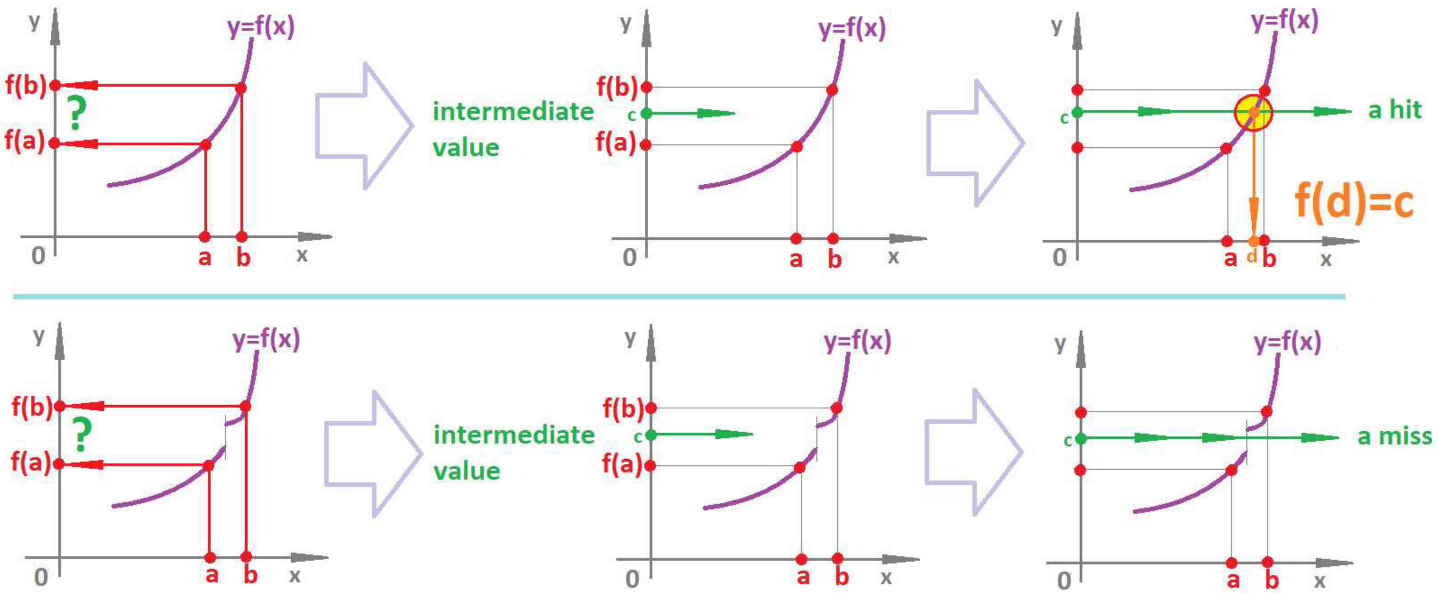
The definition of continuity is purely *local* as only the behavior of the function in the vicinity, no matter how small, of the point matters. Now, what if the function is continuous on a whole interval? What can we say about its *global* behavior?



We know what we *want* to say:

- The graph of a continuous function consists of a single piece.

Our understanding of continuity of functions has been as the property of having *no gaps in their graphs*. In fact, there are no gaps in the range either. To get from point $A = f(a)$ to point $B = f(b)$, we have to visit every point in between (no teleportation!):



We can see in the second part of the illustration how this property may fail.

This idea is more precisely expressed by the following:

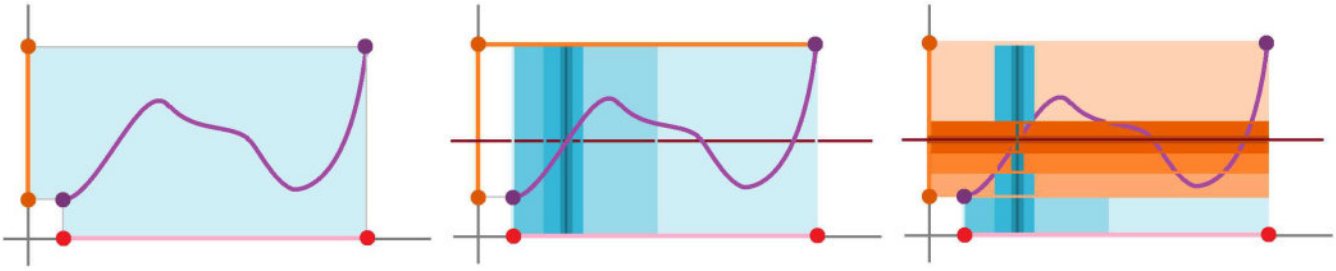
Theorem 2.9.1: Intermediate Value Theorem

A continuous function on a closed bounded interval doesn't miss any points between its values at the ends of the interval.

In other words, if a function f is defined and is continuous on an interval $[a, b]$, then for any c between $f(a)$ and $f(b)$, there is d in $[a, b]$ such that $f(d) = c$.

Proof.

The idea of the proof is this: We will make the intervals in the x-axis that satisfy the condition of the theorem narrower and narrower; then the continuity of the function will ensure that the corresponding intervals in the y-axis will get narrower and narrower too:



More precisely, we will construct such a sequence of *nested intervals*,

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset \dots ,$$

that they will have only one point in common.

Consider the halves of $I = [a, b]$:

$$\left[a, \frac{a+b}{2}\right], \ \left[\frac{a+b}{2}, b\right].$$

For at least one of them, there is a change of $y = f(x)$ from less to more or from more to less of c. Call this interval $I_1 = [a_1, b_1]$:

$$f(a_1) < c, \ f(b_1) > c \quad \text{or} \quad f(a_1) > c, \ f(b_1) < c\,.$$

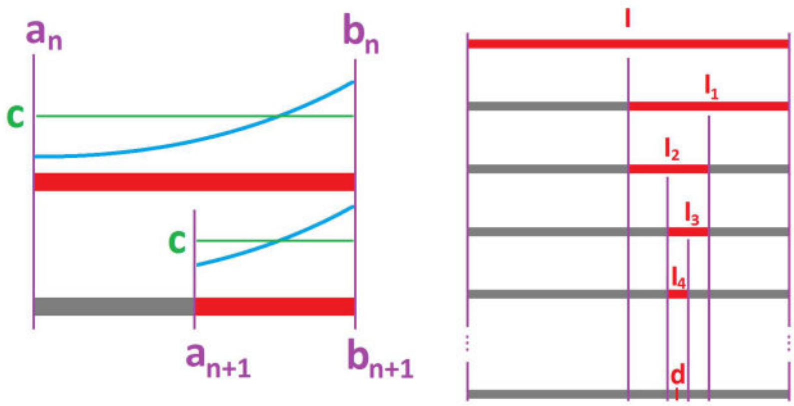
Next, we consider the halves of this new interval:

$$\left[a_1, \frac{a_1+b_1}{2}\right], \ \left[\frac{a_1+b_1}{2}, b_1\right].$$

Once again, for at least one of them, the values of f cross $y = c$. Call this interval $I_2 := [a_2, b_2]$:

$$f(a_2) < c, \ f(b_2) > c \quad \text{or} \quad f(a_2) > c, \ f(b_2) < c\,.$$

We continue:



Note that whenever $f(a_n) = c$ or $f(b_n) = c$, we are done.

Following this process, the result is a sequence of intervals:

$$I = [a, b] \supset I_1 = [a_1, b_1] \supset I_2 = [a_2, b_2] \supset \dots$$

They satisfy these two properties:

$$a \le a_1 \le \dots \le a_n \le \dots \le b_n \le \dots \le b_1 \le b\,,$$

and

$$f(a_n) < c, \ f(b_n) > c \quad \text{or} \quad f(a_n) > c, \ f(b_n) < c\,.$$

We also have:

$$|b_{n+1} - a_{n+1}| = \frac{1}{2}|b_n - a_n| = \frac{1}{2^n}|b - a| \to 0\,.$$

By the *Nested Intervals Theorem* (Chapter 1), the sequences converge, and converge to the same value:

$$a_n \to d,\ b_n \to d\,.$$

From the continuity of f, we then conclude:

$$f(a_n) \to f(d),\ f(b_n) \to f(d)\,.$$

By the *Comparison Theorem*, we conclude from the inequalities above:

$$f(d) \le c,\ f(d) \ge c \quad \text{or} \quad f(d) \ge c,\ f(d) \le c\,.$$

Hence, we have:

$$f(d) = c\,.$$

So, d is the desired number.

Choosing $c = 0$ in the theorem gives us the following:

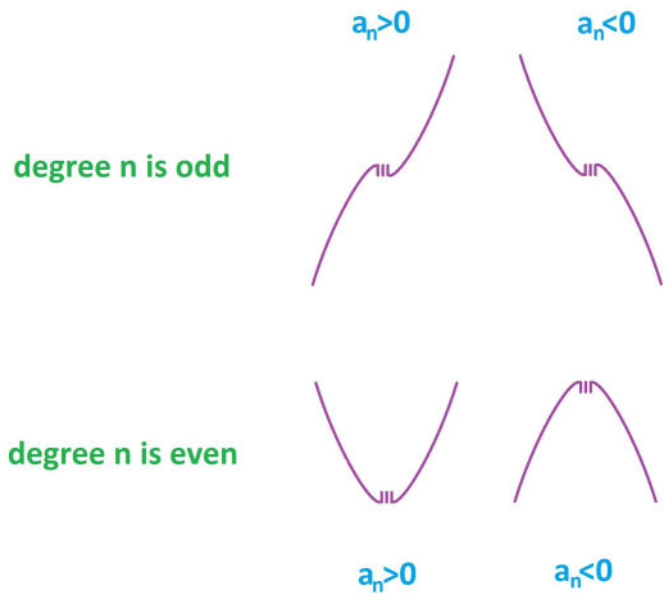
Corollary 2.9.2: x-intercepts of Continuous Function

If a continuous on a closed interval function takes values with opposite signs at its end-points, then it has an x-intercept within this interval.

In other words, if f is continuous on $[a, b]$, then we have:

$$f(a) > 0,\ f(b) < 0 \ \ \texttt{OR} \ \ f(a) < 0,\ f(b) > 0 \implies f(c) = 0 \ \textit{ for some } c \ \textit{ in } \ [a, b]\,.$$

We can make this observation even more specific, for polynomials. Recall how they are classified:



Indeed, they have infinite – but possibly different – limits at the infinities. The ones in the first row are, therefore, guaranteed to take both positive and negative values. We conclude the following:

Corollary 2.9.3: x-intercepts of Polynomial

Any odd degree polynomial has an x-intercept.

Exercise 2.9.4

Provide details to the proof of the corollary.

Example 2.9.5: iterated search

The proof of the *Intermediate Value Theorem* via the *Nested Intervals Theorem* is nothing more than an *iterated search for a solution of the equation* $f(x) = c$. Let's use the method to solve this equation:

$$\sin x = 0 .$$

We start with the interval $[a_1, b_1] = [3, 3.5]$. The function $f(x) = \sin x$ does change its sign over this interval. We divide the interval in half and pick the half over which f also changes its sign. Then we repeat this process several times. The spreadsheet formula for a_n is:

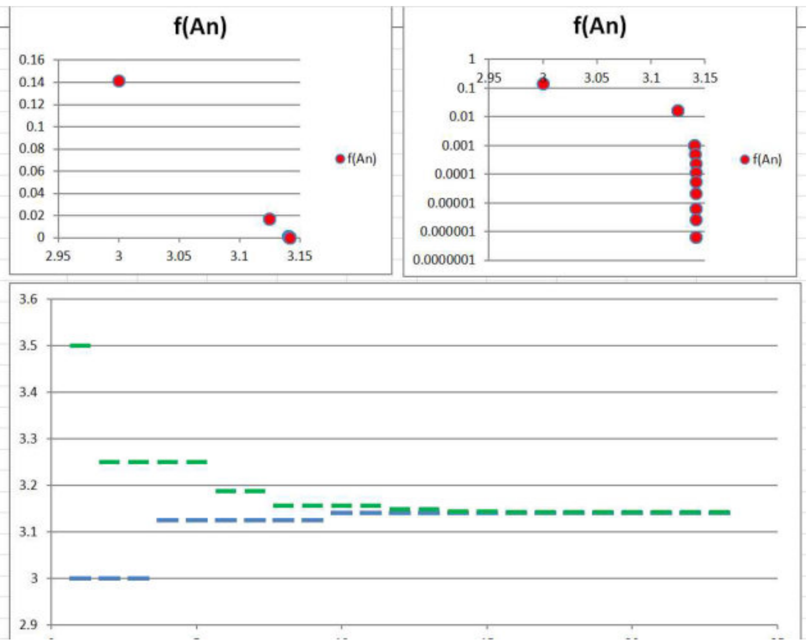
```
=IF(R[-1]C[3]*R[-1]C[4]<0,R[-1]C,R[-1]C[1])
```

and b_n:

```
=IF(R[-1]C[1]*R[-1]C[2]<0,R[-1]C[-1],R[-1]C)
```

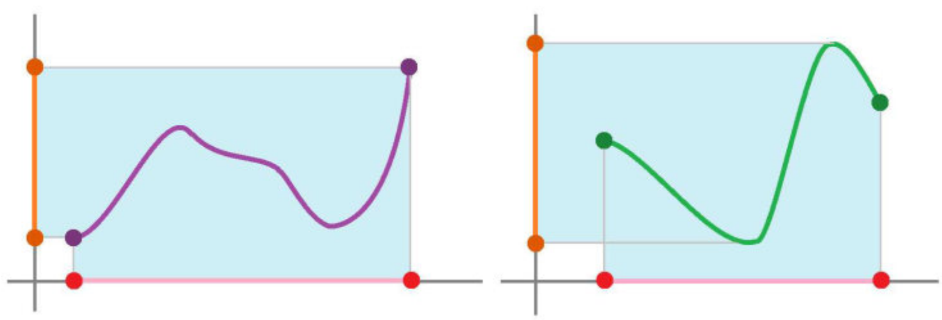
while d_n is the mid-point of the interval:

n	An	Dn	Bn	f(An)	f(Dn)	f(Bn)	+/-
0	3	3.25	3.5	0.14112	-0.1082	-0.35078	-
1	3	3.25	3.25	0.14112	-0.1082	-0.1082	-
2	3	3.125	3.25	0.14112	0.016592	-0.1082	-
3	3.125	3.125	3.25	0.016592	0.016592	-0.1082	-
4	3.125	3.1875	3.25	0.016592	-0.04589	-0.1082	-
5	3.125	3.1875	3.1875	0.016592	-0.04589	-0.04589	-
6	3.125	3.15625	3.1875	0.016592	-0.01466	-0.04589	-
7	3.125	3.15625	3.15625	0.016592	-0.01466	-0.01466	-
8	3.125	3.140625	3.15625	0.016592	0.000968	-0.01466	-
9	3.140625	3.140625	3.15625	0.000968	0.000968	-0.01466	-
10	3.140625	3.148438	3.15625	0.000968	-0.00684	-0.01466	-
11	3.140625	3.148438	3.148438	0.000968	-0.00684	-0.00684	-
12	3.140625	3.144531	3.148438	0.000968	-0.00294	-0.00684	-
13	3.140625	3.144531	3.144531	0.000968	-0.00294	-0.00294	-
14	3.140625	3.142578	3.144531	0.000968	-0.00099	-0.00294	-
15	3.140625	3.142578	3.142578	0.000968	-0.00099	-0.00099	-
16	3.140625	3.141602	3.142578	0.000968	-8.9E-06	-0.00099	-
17	3.140625	3.141602	3.141602	0.000968	-8.9E-06	-8.9E-06	-
18	3.140625	3.141113	3.141602	0.000968	0.000479	-8.9E-06	-
19	3.141113	3.141113	3.141602	0.000479	0.000479	-8.9E-06	-
20	3.141113	3.141357	3.141602	0.000479	0.000235	-8.9E-06	-
21	3.141357	3.141357	3.141602	0.000235	0.000235	-8.9E-06	-
22	3.141357	3.141479	3.141602	0.000235	0.000113	-8.9E-06	-
23	3.141479	3.141479	3.141602	0.000113	0.000113	-8.9E-06	-
24	3.141479	3.141541	3.141602	0.000113	5.21E-05	-8.9E-06	-
25	3.141541	3.141541	3.141602	5.21E-05	5.21E-05	-8.9E-06	-
26	3.141541	3.141571	3.141602	5.21E-05	2.16E-05	-8.9E-06	-
27	3.141571	3.141571	3.141602	2.16E-05	2.16E-05	-8.9E-06	-
28	3.141571	3.141586	3.141602	2.16E-05	6.35E-06	-8.9E-06	-



We can see that the values of a_n, b_n converge to π and the values of $f(a_n)$, $f(b_n)$ to 0. The scatter plot of $(a_n, f(a_n))$ is shown on the right along with its version for a logarithmically re-scaled y-axis.

The theorem says that there are no missing values in the image of an interval. It's also an interval:



More concisely, we have the following:

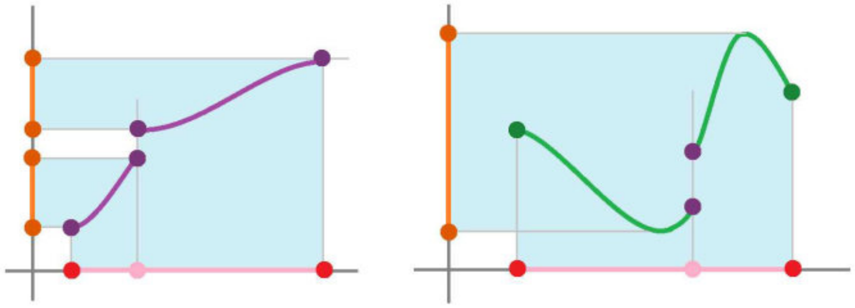
Corollary 2.9.6: Range of Continuous Function

If the domain of a continuous function is an interval, then so is its range.

Proof.

It follows from the *Intermediate Point Theorem*.

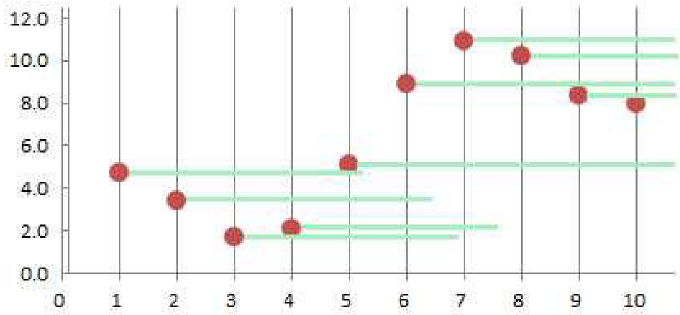
This is how discontinuity may cause the range to have gaps:



The second graph shows that the converse of the theorem isn't true.

Recall (from Volume 1, Chapter 1PC-4) that all strictly monotonic functions are one-to-one because the graph can't come back and cross a horizontal line for the second time:

The converse is true but only for continuous functions:



Exercise 2.9.7

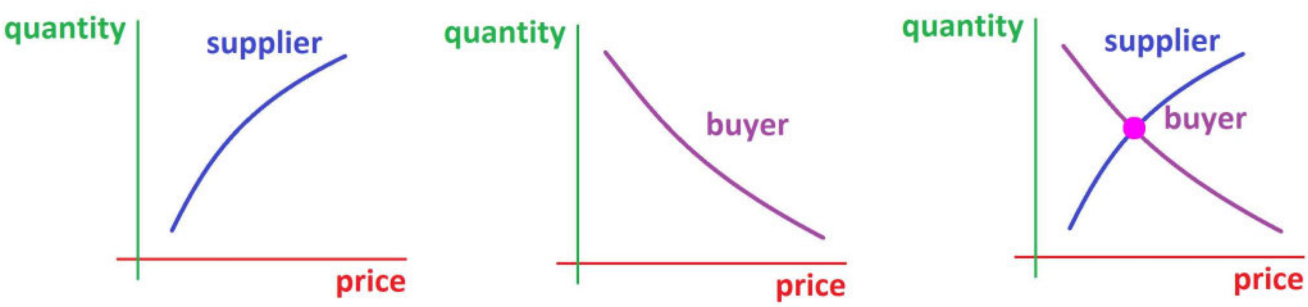
Prove that if we have a continuous function defined on an interval I, then the function is one-to-one if and only if it is strictly monotonic.

Example 2.9.8: supply and demand

Here is a continuity argument used in economics. In a typical transaction, we have the following:

- The supplier is willing to produce more of the commodity for a higher price.
- The buyer is prepared to buy more for a lower price.

Either fact implies that there is a function: how the quantity of the commodity depends on price. The former is increasing and the latter is decreasing:



If we assume also that the two functions are *continuous*, we conclude that there must be a price that satisfies both parties.

Exercise 2.9.9

What other implicit assumptions are made in the example? Put the example in the form of a theorem and prove it.

So, the image of an interval is an interval, but is the image of a *closed* interval closed? But first, is the function bounded?

Recall that a function f is called *bounded* on an interval I if its range is bounded, i.e., there is such a real number Q that

$$|f(x)| \leq Q$$

for all x in I. The link to the current subject is the result below:

Theorem 2.9.10: Convergent Means Bounded

If the limit of a function exists, then the function is bounded on some open interval that contains this point.

In other words, we have:

$$\lim_{x\to a} f(x) \text{ exists} \implies |f(x)| \leq Q$$

for all x in $(a-\delta, a+\delta)$ for some $\delta > 0$ and some Q.

Exercise 2.9.11

Prove the theorem.

We have been speaking until now of continuity only one point at a time; there is no cut at $x = a$. Now *sets*:

Definition 2.9.12: function continuous on interval

A function f is called *continuous on an interval* I if the interval is contained in the domain of f and

$$\lim_{n\to\infty} f(x_n) = f(a)$$

for any sequence x_n in I such that $x_n \to a$.

Example 2.9.13: continuity on intervals

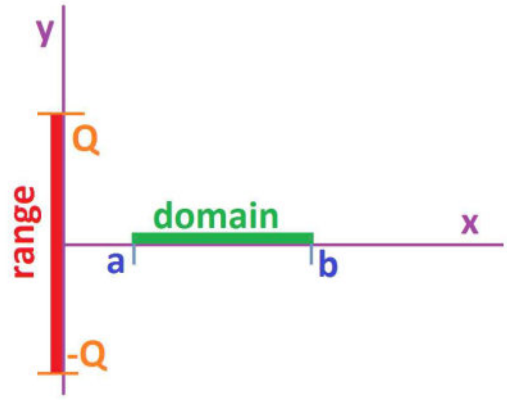
In particular, we have:

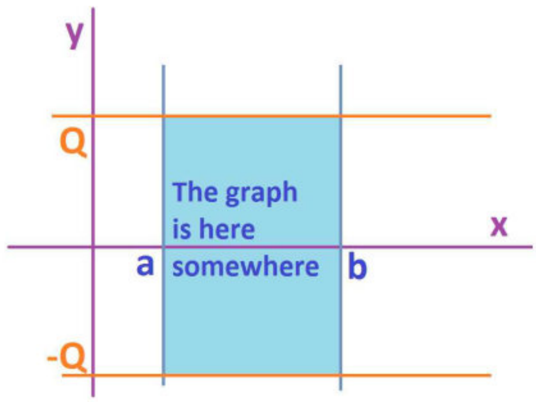
- The function $1/x$ is continuous on $(-\infty, 0)$ and on $(0, \infty)$.
- But it is not continuous on $(-\infty, \infty)$ or even $(-\infty, 0) \cup (0, \infty)$.

We also have:

- The function $\sqrt{x}$ is continuous on $[0, \infty)$.
- The functions $\sin^{-1} x$ and $\cos^{-1} x$ are continuous on $[-1, 1]$.

The *global* version of the above theorem guarantees that the function is bounded on any closed bounded interval, i.e., $[a, b]$, as follows:





In other words, we have:

Theorem 2.9.14: Continuous Means Bounded

A function continuous on a closed bounded interval is bounded on the interval.

Proof.

Suppose, to the contrary, that f is unbounded on interval $[a, b]$. Then there is a sequence x_n in $[a, b]$ such that $f(x_n) \to \infty$. Then, by the *Bolzano-Weierstrass Theorem* (Chapter 1), the sequence x_n has a convergent subsequence y_k:

$$y_k \to y\,.$$

This point belongs to $[a, b]$! From the continuity of the function, it follows that

$$f(y_k) \to f(y)\,.$$

This contradicts the fact that y_k is a subsequence of a sequence that diverges to ∞.

In other words, the *range* or the image of the function is within $[-Q, Q]$.

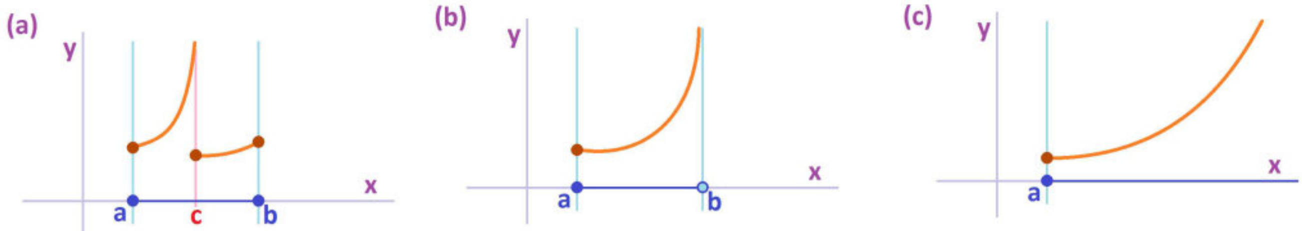
Exercise 2.9.15

Why are we justified to conclude in the proof that the limit y of y_k is in $[a, b]$?

Example 2.9.16: pre-conditions

We show below that the theorem fails if one of the conditions is omitted:

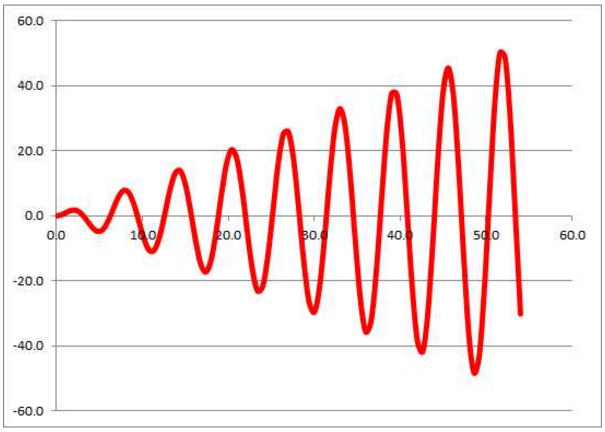
1. The function is *discontinuous* on a closed bounded interval.
2. The function is continuous on a *not closed* bounded interval.
3. The function is continuous on a closed *unbounded* interval.



In all three cases, the unbounded behavior of the function can be expressed as an infinite limit:

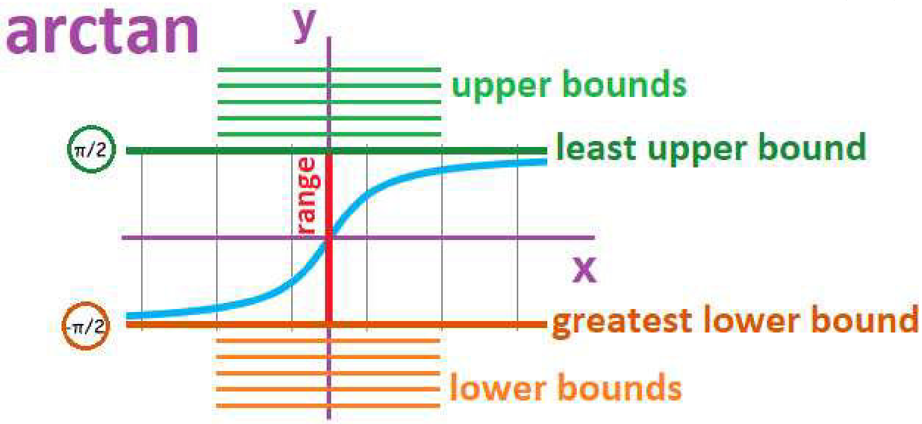
$$\begin{array}{ll} 1. & \lim\limits_{x \to c^-} f(x) = +\infty \\ 2. & \lim\limits_{x \to b^-} f(x) = +\infty \\ 3. & \lim\limits_{x \to +\infty} f(x) = +\infty \end{array}$$

However, we shouldn't equate unboundedness with infinite limits; here is the graph of $y = x \sin x$:



Example 2.9.17: arctan

The range may be an open interval, as in the case of arctangent. Its range is $(-\pi/2, \pi/2)$:



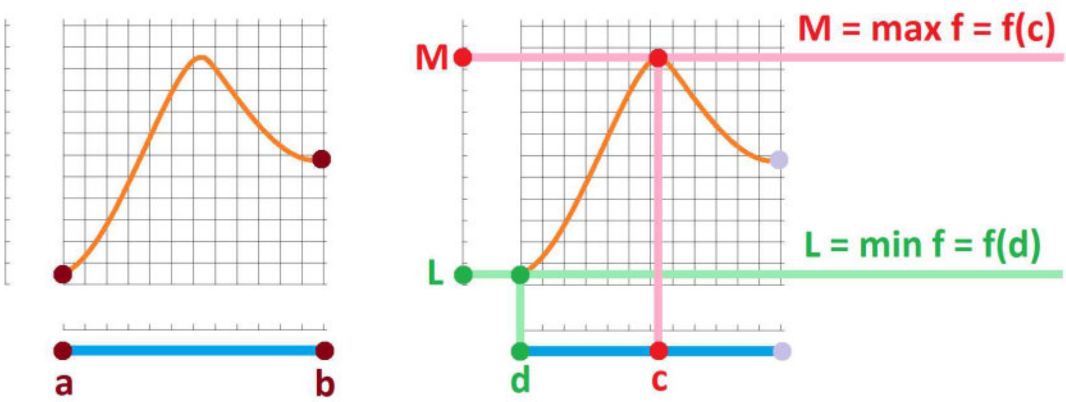
Of course we have:

$$\inf(-\pi/2, \pi/2) = -\pi/2, \ \sup(-\pi/2, \pi/2) = \pi/2 .$$

However, these values are never reached:

$$\arctan(x) \neq \pm\pi/2 .$$

If the image/range is a closed interval, the function *reaches its extreme values*, i.e., the maximum and minimum of its range.



Let's recall the relevant concepts (from Volume 1, Chapter 1PC-4):

Definition 2.9.18: global maximum and minimum points

Suppose we have a function $y = f(x)$ and the interval $[a, b]$ is within its domain.

- A point $x = d$ is called a *global maximum point* of f on interval $[a, b]$ if

$$f(d) \geq f(x) \text{ for all } a \leq x \leq b .$$

- A point $x = c$ is called a *global minimum point* of f on interval $[a, b]$ if

$$f(c) \leq f(x) \text{ for all } a \leq x \leq b\,.$$

Collectively they are all called *global extreme points*. The word "global" is often omitted.

Just because something is described doesn't mean that it can be found. For example, $f(x) = 1/x$ has no minimum value on $[0, \infty)$. The result below guarantees their existence:

Theorem 2.9.19: Extreme Value Theorem

A continuous function on a bounded closed interval has a global maximum and a global minimum.

In other words, if f is continuous on $[a, b]$, then there are c, d in $[a, b]$ such that

$$f(c) \leq f(x) \leq f(d)$$

for all x in $[a, b]$.

Proof.

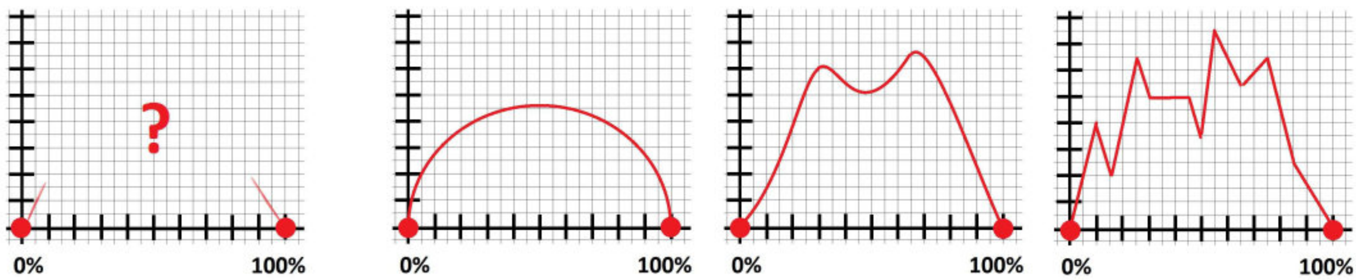
According to the last two theorems, the image of $[a, b]$ is a bounded interval.

Exercise 2.9.20

Provide the rest of the proof. Hint: the *Bolzano-Weierstrass Theorem* (Chapter 1).

Example 2.9.21: Laffer curve

Here is continuity argument used in economics. The exact dependence of the revenue on the income tax rate is unknown. It seems obvious, however, that both 0% and 100% rates will produce zero revenue.



If we also assume that the dependence is *continuous*, then there is a maximum point somewhere within the interval. We conclude that increasing the rate may yield a tax-decreasing revenue.

Exercise 2.9.22

What other implicit assumptions are made in the example? Put the example in the form of a theorem and prove it.

More relevant concepts below:

Definition 2.9.23: global maximum and minimum values

Suppose we have a function $y = f(x)$ and the interval $[a, b]$ is within its domain.

- A point $y = M$ is called the *global maximum value* of f on interval $[a, b]$ if

it is the maximum element of the range of f; i.e.,

$$M = \max\{f(x) : \ a \le x \le b\}\,.$$

- A point $y = m$ is called the *global minimum value* of f on interval $[a, b]$ if it is the minimum element of the range of f; i.e.,

$$m = \min\{f(x) : \ a \le x \le b\}\,.$$

Collectively they are all called *global extreme values*. The word "global" is often omitted.

This means simply that we have:

$$f\big(\text{ maximum point }\big) = \text{ maximum value } \quad \text{and} \quad f\big(\text{ minimum point }\big) = \text{ minimum value}$$

Warning!

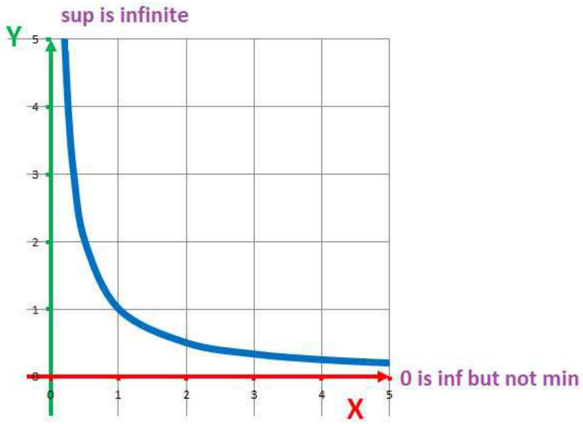
The absolute maximum and minimum values and points are also called *absolute maximum and minimum* values and points.

Then *the* global max (or min) value is reached by the function at *any* of its global max (or min) points. For example, $f(x) = \sin x$ attains its max value of 1 for infinitely many choices of $x = \pi/2, 5\pi/2, \dots$.

Why are the restrictions of theorem essential?

Example 2.9.24: $1/x$

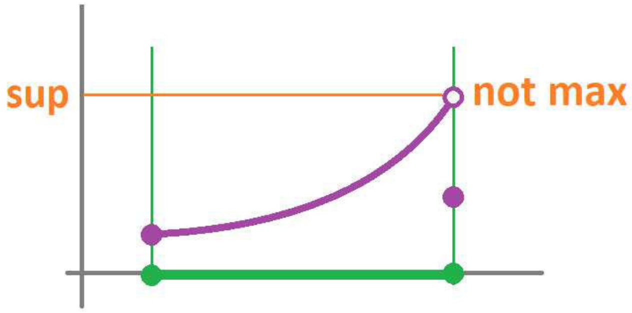
The function $1/x$ doesn't attain its least upper bound value on $(0, 1]$ as it is, in fact, infinite:



The theorem doesn't apply because the interval *isn't closed*. Secondly, the function $1/x$ doesn't attain its greatest lower bound value, which is 0, on $[1, \infty)$. The theorem doesn't apply because the interval *isn't bounded*.

Example 2.9.25: discontinuous

Here the maximum isn't attained even though the interval is closed and bounded:

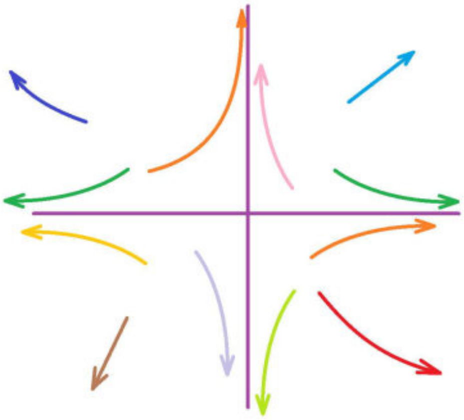


The theorem doesn't apply because f *isn't continuous*.

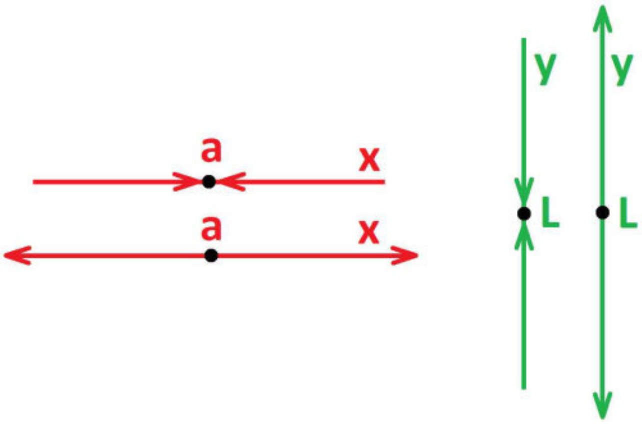
Note that the reason we need the *Extreme Value Theorem* is to ensure that the optimization problem we are facing has a solution.

2.10. Large-scale behavior and asymptotes

Graphs of most functions are infinite in size and won't fit on any piece of paper. They have to *leave* the paper and they do that in a number of different ways:



In order to describe this behavior algebraically, we will look at the x- and y-coordinate of the points of the graph separately. For either, there are two main cases:



There can be either:

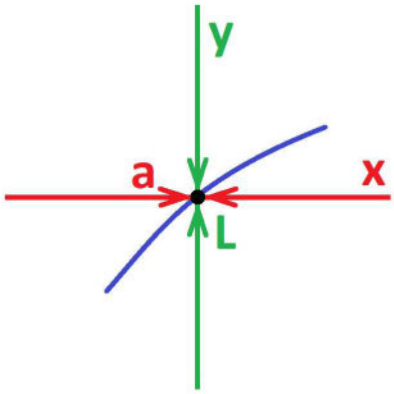
- a finite limit (from one side) or
- an infinite limit (positive or negative).

The coordinates are independent, and the two cases can be combined in a number of ways.

First, both limits are finite:

$$x \to a^{\pm},\ y \to L^{\pm}\,.$$

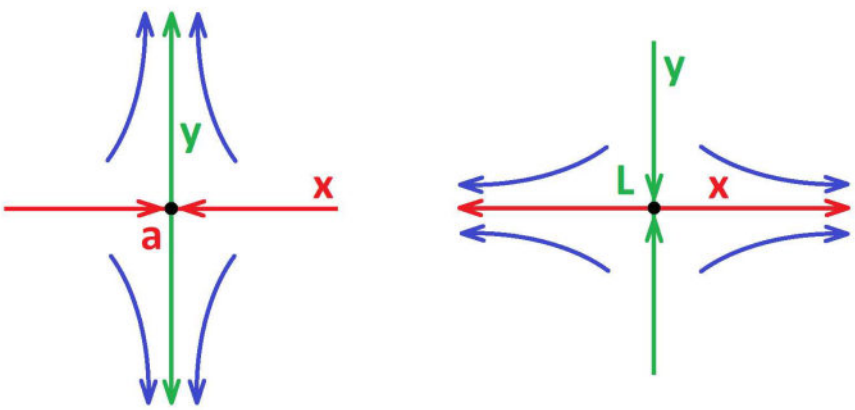
This is the familiar limit that has been discussed from the very beginning of the chapter:



Second, one limit is finite and the other infinite:

$$x \to a^{\pm},\ y \to \pm\infty \quad \text{or} \quad x \to \pm\infty\ y \to L^{\pm}\,.$$

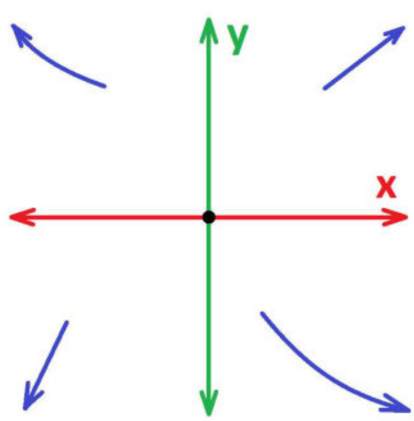
There are eight possibilities here, all with either horizontal or vertical asymptotes:



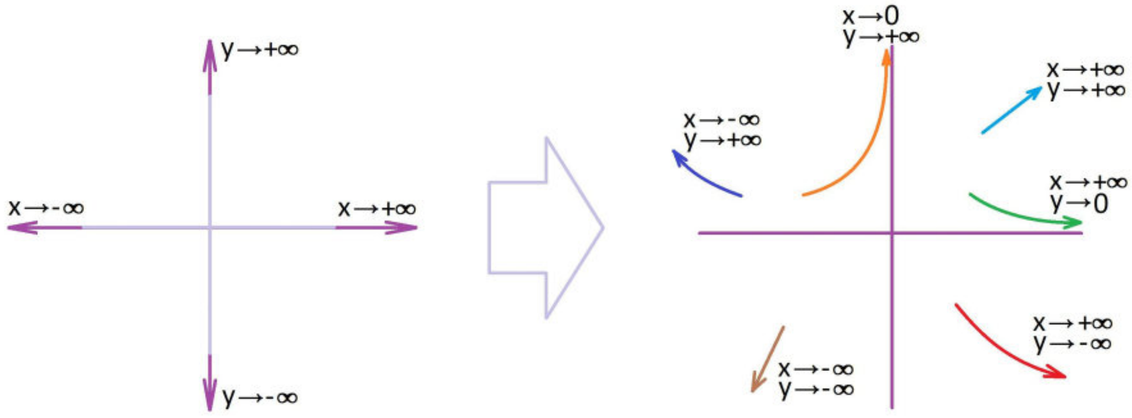
Third, both limits are infinite:

$$x \to \pm\infty,\ y \to \pm\infty\,.$$

There are four possibilities, all with no asymptotes:



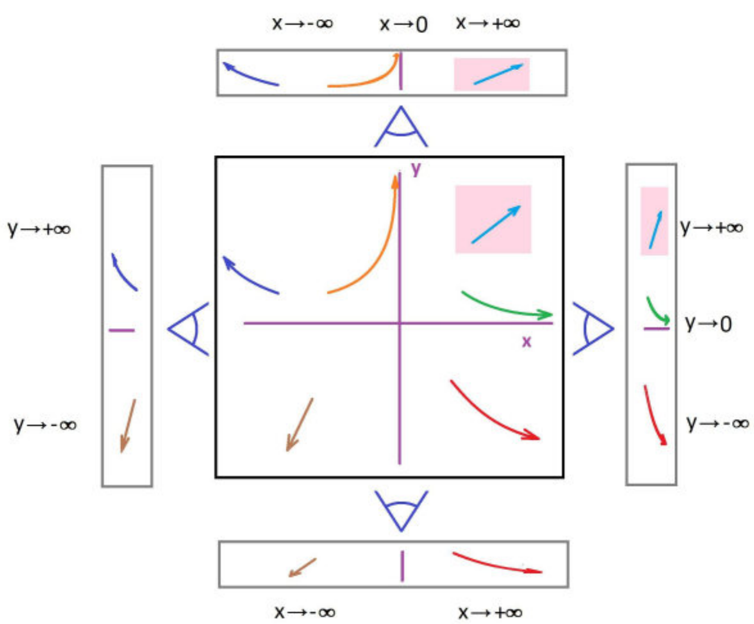
In order to determine the trend of the point (x, y) in the plane, we can imagine walking *on* the curve and looking down on the x-axis to record the x-coordinate and looking forward or back to see what is happening the y-coordinate:



Exercise 2.10.1

Suggest a function for each of these types of trends.

Alternatively, we imagine that the curves are drawn on a piece of paper. Then, if we look at it at a sharp angle from the direction of the x-axis, the change of x becomes almost invisible and we clearly see the behavior of the y-coordinate:

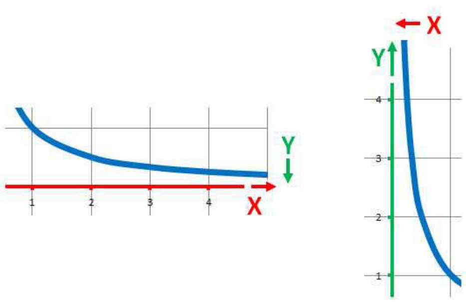


If we look from the direction of the y-axis, only the change of x is significant.

Conversely, if we have the limit description of the curve, what does it look like? For example, we might have:

- $x \to +\infty$ and $y \to 0^+$
- $x \to 0^+$ and $y \to -\infty$

If the point approaches a line it can't cross ($y = 0$ and $x = 0$), the curve starts to become more and more straight and slowly approaches that line:



Eventually it *almost* merges with this line. The line is then called an *asymptote*.

Example 2.10.2: no asymptotes

The simplest case, however, is the one with no asymptotes:

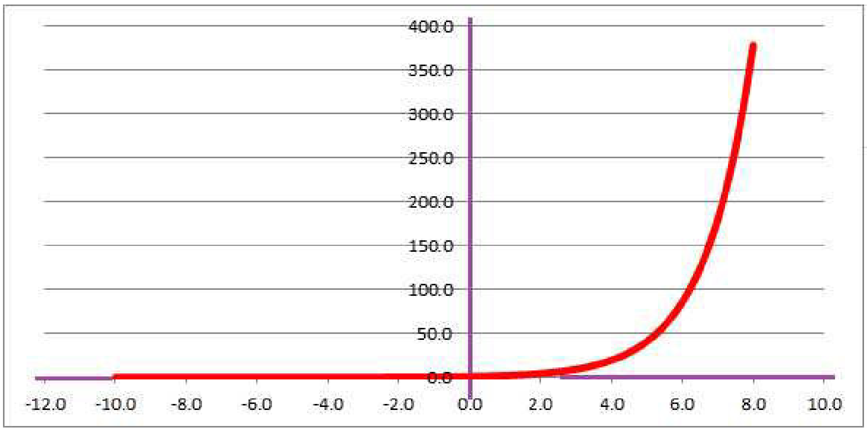
$$y = x,\ y = x^2, \dots \text{ at } \pm\infty,\ y = e^x,\ y = \ln x \text{ at } +\infty\,,$$

and many more.

Example 2.10.3: exponential

We previously demonstrated the following about the exponential function:

$$x \to -\infty,\ y \to 0 \text{ and } x \to +\infty,\ y \to +\infty\,.$$

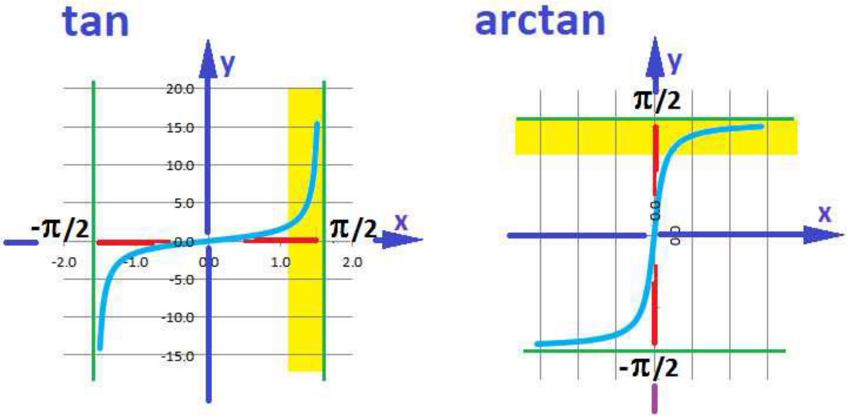


Example 2.10.4: arctangent

As an illustration of the asymptotic behavior, we will consider the (restricted) tangent $y = \tan x$ and its inverse, arctangent:

$$y = \tan x \text{ and } x = \tan^{-1} y$$

for $-\pi/2 < x < \pi/2$. The graphs are, of course, the same with just x and y interchanged. Let's describe their large-scale behaviors at a single area (in yellow) using limits:



First the tangent:

$$y \to +\infty \text{ as } x \to \pi/2^-\,.$$

Changing x to be the dependent and y to be the independent variables, we simply rewrite the above for the arctangent:

$$y \to \pi/2^- \text{ as } x \to +\infty\,.$$

Warning!

Describing asymptotic behavior as "flattening" of the graph is misleading, as the example $y = \sqrt{x}$ shows.

The three patterns of behavior of a function as its graph leaves the visible part of the xy-plane are stated below in precise terms:

Definition 2.10.5: function goes to infinity

Given a function f, we say that f *goes to infinity on the right* if

$$f(x_n) \to \pm\infty \ \text{ for any sequence } x_n \to +\infty\,.$$

We use the notation:

$$f(x) \to \pm\infty \text{ as } x \to +\infty$$

or

$$\lim_{x\to+\infty} f(x) = \pm\infty\,.$$

Under this definition, we say f *goes to infinity on the left* when $x \to -\infty$.

Definition 2.10.6: horizontal asymptote as limit

Given a function $y = f(x)$, a line $y = p$ for some real p is called a *horizontal asymptote* of f if

$$\lim_{n\to\infty} f(x_n) = p \ \text{ for any sequence } x_n \to +\infty\,.$$

We use the notation:

$$f(x) \to p \text{ as } x \to +\infty$$

or, alternatively,

$$\lim_{x\to+\infty} f(x) = p\,.$$

Under this definition, we can also have $x \to -\infty$.

Definition 2.10.7: vertical asymptote as limit

A line $x = a$ for some real a is called a *vertical asymptote* of f if

$$\lim_{n\to\infty} f(x) = \pm\infty \ \text{ for any sequence } x_n \to a^+\,.$$

We use the notation:

$$f(x) \to \pm\infty \text{ as } x \to a^+$$

or, alternatively,

$$\lim_{x\to a^+} f(x) = \pm\infty\,.$$

Under this definition, we can also have $x \to a^-$.

The fourth pattern is "no pattern".

Example 2.10.8: notation

We collect this information for the exponential function:

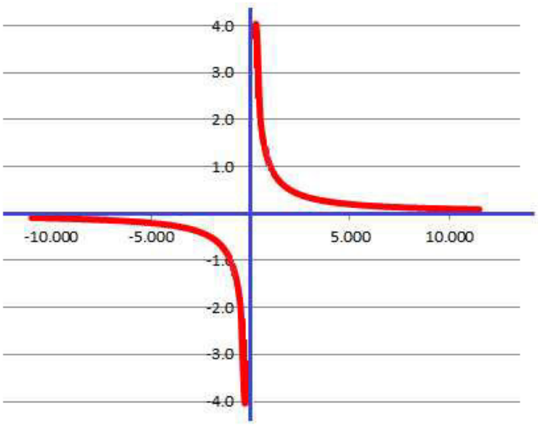
$$\lim_{x\to-\infty} e^x = 0, \quad \lim_{x\to+\infty} e^x = +\infty\,,$$

and for the arctangent:

$$\lim_{x\to-\infty} \tan^{-1} = -\pi/2^+ \text{ and } \lim_{x\to+\infty} \tan^{-1} y \to \pi/2^-\,.$$

Example 2.10.9: symmetry

Even though the last two definitions look very different, they describe the *identical* behavior of the curve.



We can see this symmetry in the graph of $y = 1/x$, and we can see it in the algebra:

	x	y
horizontal:	$x \to +\infty$	$y \to 0$
vertical:	$x \to 0$	$y \to +\infty$

We just interchange x and y. In more detail:

$$\frac{1}{x} \to 0^- \text{ as } x \to -\infty,\ \frac{1}{x} \to 0^+ \text{ as } x \to +\infty, \frac{1}{x} \to -\infty \text{ as } x \to 0^-,\ \frac{1}{x} \to +\infty \text{ as } x \to 0^+\,.$$

The analysis suggests the following:

Theorem 2.10.10: Asymptotes of Inverse

The vertical asymptotes of a function are the horizontal asymptotes of its inverse, and vice versa.

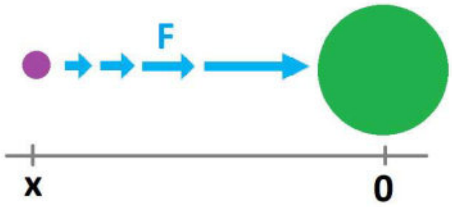
Exercise 2.10.11

Prove the theorem.

Example 2.10.12: Newton's Law of Gravity

The law states:

► The force of gravity between two objects is inversely proportional to the square of the distance between their centers.



In other words, the force is given by the formula:

$$F = \frac{C}{r^2},$$

where:
- C is some constant, and
- r is the distance between the centers of the mass of the two.

Then, if $r > 0$ is variable, we have:

$$\lim_{r\to+\infty} F(r) = 0.$$

This means that the force becomes negligible when the two objects become sufficiently far away from each other.

Exercise 2.10.13

What is happening when the two bodies are getting closer and closer to each other?

Example 2.10.14: Newton's Law of Cooling

We derived from the law the following fact:

▶ The difference between an object's temperature and the temperature of the atmosphere is declining exponentially.

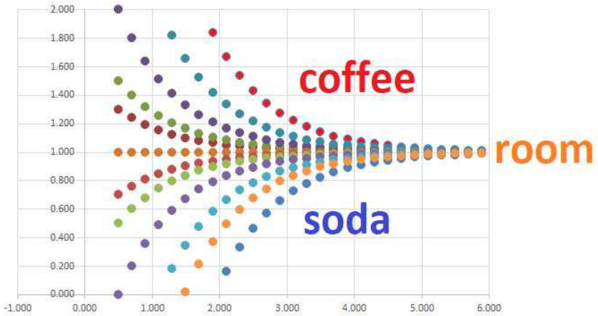
In other words, we have:

$$T - T_0 = Ce^{kx},\ k < 0,\ C > 0,$$

where T is the current temperature and T_0 is the ambient temperature. We compute:

$$\lim_{t\to+\infty} T(t) = T_0.$$

Therefore, there is a horizontal asymptote $y = T_0$. The asymptote may be approached by the graph from above or from below:
- A warmer object is cooling: $T - T_0 > 0$ and $T\searrow$.
- A cooler object is warming: $T - T_0 < 0$ and $T\nearrow$.



Example 2.10.15: plots from limits

Suppose we are facing the opposite (inverse!) problem: Suppose we know the limits and now we need to plot the asymptotes and a possible graph of the function.

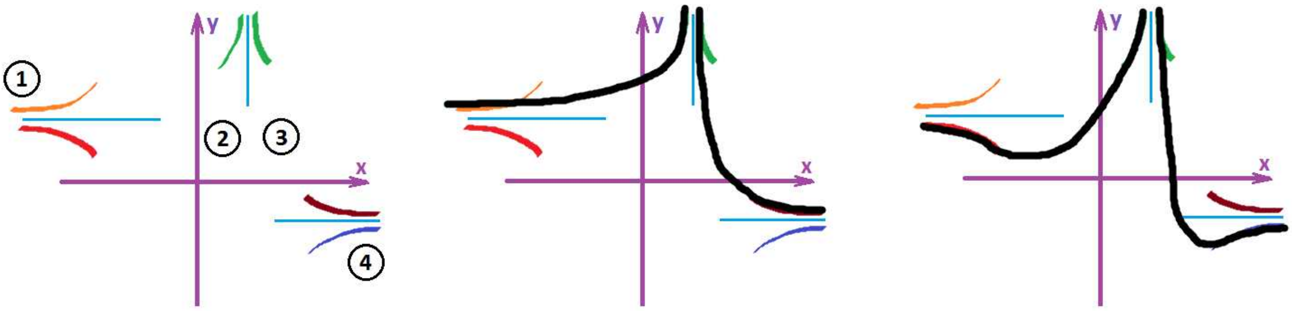
Suppose this is what we know about f: It is defined for all $x \neq 3$, and

$$\lim_{x\to-\infty} f(x) = 3,\ \lim_{x\to 3} f(x) = +\infty,\ \lim_{x\to+\infty} f(x) = -2\,.$$

Let's rewrite those:

1. $x \to -\infty,\ y \to 3$
2. $x \to 3^-,\ y \to +\infty$
3. $x \to 3^+,\ y \to +\infty$
4. $x \to +\infty,\ y \to -2$

We draw rough strokes to represent these facts (left):



The ambiguity about how the graph approaches the asymptotes remains at $-\infty$ and $+\infty$. We connect the initial strokes into a single graph; it has two branches. Two possible versions of the graph of f are shown (middle and right).

Example 2.10.16: three patterns

A horizontal asymptote may be approached by the graph is these three main ways:

They are exemplified by the following limits:

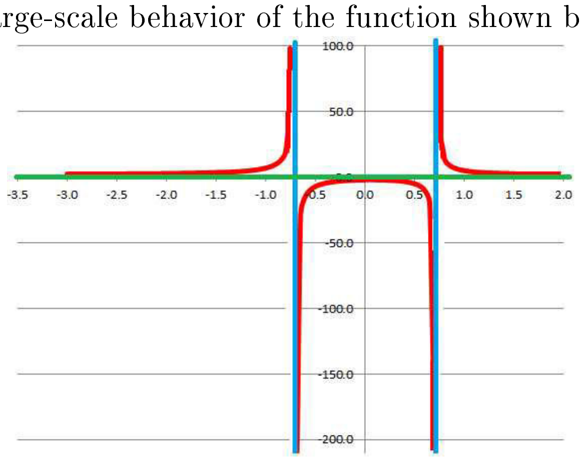
$$\lim_{x\to\infty}\left(1+\frac{1}{x}\right)=1,\quad \lim_{x\to+\infty}\left(2-\frac{1}{x}\right)=2,\quad \lim_{x\to+\infty}\left(\frac{1}{x}\cos x\right)=0\,.$$

Exercise 2.10.17

Suggest a function that oscillates like the last one but doesn't flatten out.

Exercise 2.10.18

Use limits to describe the large-scale behavior of the function shown below:



2.11. Limits and infinity: computations

Let's summarize the possible outcomes of our computations of a limit (at a point, one-sided or two-sided, or at infinity):

$$\lim = \begin{cases} L & \to \text{ It's a number. You can do algebra with the limit.} \\ \pm\infty & \to \text{ You can do some algebra: } \infty+\infty=\infty\text{, but not all: } \infty-\infty=? \\ \text{no limit} & \to \text{ Do no algebra with the limit.} \\ 0/0,\ \infty/\infty,\ \infty-\infty & \to \text{ It's indeterminate. Start over!} \end{cases}$$

Example 2.11.1: complete analysis

Let's fully investigate this function:

$$f(x) = \begin{cases} \dfrac{x^2-x}{x^2-1} & \text{if } x \neq 1\,, \\ 1 & \text{if } x = 1\,. \end{cases}$$

The point where the two formulas might conflict is $x = 1$. The function is defined, but does the limit exist? Let's compute:

$$\lim_{x\to 1} f(x) = \lim_{x\to 1} \frac{x^2-x}{x^2-1}\,.$$

Plug in $x = 1$ (mistake!):

$$\frac{1^1-1}{1^2-1} = \frac{0}{0}\,. \quad \boxed{\textbf{DEAD END}}$$

Indeterminate! We conclude that we need to do algebra with the function before turning back to the limit.

Here is the algebra:

$$\begin{aligned} \lim_{x\to 1} \frac{x^2-x}{x^2-1} &= \lim_{x\to 1} \frac{x(x-1)}{(x-1)(x+1)} \\ &= \lim_{x\to 1} \frac{x}{x+1} \\ &= \frac{1}{1+1} = \frac{1}{2}\,. \end{aligned}$$

The last step (plugging in $x = 1$) is justified by the fact that the function $\dfrac{x}{x+1}$ is rational and, therefore, continuous at any point of its domain. Now, are the two one-sided limits equal? No: $1 \neq \dfrac{1}{2}$. That's why the function has a discontinuity at this point (a removable kind).

We can simplify the formula:

$$f(x) = \begin{cases} \dfrac{x}{x+1} & \text{if } x \neq 1\,, \\ 1 & \text{if } x = 1\,. \end{cases}$$

We find the domain by solving: $x + 1 \neq 0$ and $x \neq 1$. The domain is $(-\infty, -1) \cup (-1, \infty)\,.$

Consider $\dfrac{x}{x+1}$ at $x = -1$:

$$\lim_{x\to -1} \frac{x}{x+1} = \infty\,.$$

Thus, $x = -1$ is a vertical asymptote.

However, there is still the issue of in which of the four different ways the graph approaches the asymptote. To tell which one, compute the one-sided limits while watching the signs of the numerator and the denominator:

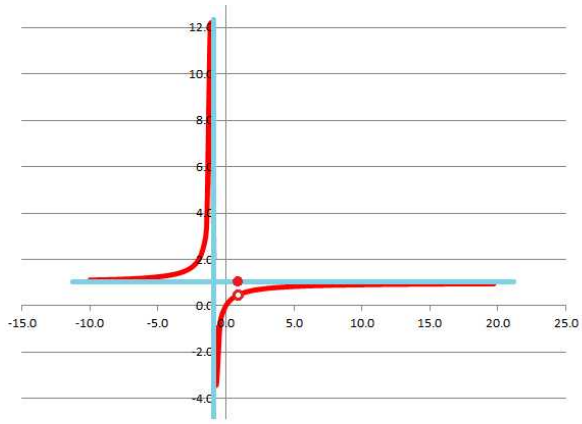
$$\lim_{x\to -1^-} \frac{x}{x+1} \left(= \frac{-}{-} = + \right) = +\infty,$$
$$\lim_{x\to -1^+} \frac{x}{x+1} \left(= \frac{-}{+} = - \right) = -\infty\,.$$

Next, the function's behavior at ∞:

$$\lim_{x\to\infty} f(x) = \lim_{x\to\infty} \frac{x}{x+1} = 1\,.$$

Thus, $y = 1$ is a horizontal asymptote.

Finally, we confirm the plot:



Just as before, each property of limits is matched by its analog for limits at infinity:

Theorem 2.11.2: Algebra of Limits of Functions at Infinity

Suppose $f(x) \to F$ and $g(x) \to G$ as $x \to +\infty$ (or $-\infty$). Then

SR:	$f(x) + g(x) \to F + G$	*CMR:*	$c \cdot f(x) \to cF$	*for any real c*
PR:	$f(x) \cdot g(x) \to FG$	*QR:*	$f(x)/g(x) \to F/G$	*when $G \neq 0$*

Theorem 2.11.3: Substitution Rule for Limits

Suppose f and g are two functions. Then, if f is continuous at $L = \lim\limits_{x\to\infty} g(x)$, then we have:

$$\boxed{\lim_{x\to\infty} f(g(x)) = f\left(\lim_{x\to\infty} g(x)\right)}$$

Example 2.11.4: substitution

Compute:

$$\begin{aligned}
\lim_{x\to-\infty}\left(e^x - \frac{1}{x} + 3\right) &\overset{\text{SR}}{=\!=} \lim_{x\to-\infty} e^x - \lim_{x\to-\infty}\frac{1}{x} + \lim_{x\to-\infty} 3 \\
&= 0 - 0 + 3 \\
&= 3\,.
\end{aligned}$$

Example 2.11.5: horizontal asymptote

Compute:

$$\lim_{x\to\infty} \frac{x}{x+1}\,.$$

Plugging in $x=\infty$ (mistake!), gives us

$$\frac{\infty}{\infty} \quad \boxed{\textbf{DEAD END}}$$

Facing an indeterminate expression, we are supposed to do algebra instead. What are we trying to accomplish? We want to get rid of the infinities in the numerator and the denominator. How about we divide both by x?

$$\begin{aligned}
\lim_{x\to\infty} \frac{x}{x+1} &= \lim_{x\to\infty} \frac{\frac{x}{x}}{\frac{(x+1)}{x}} \\
&= \lim_{x\to\infty} \frac{1}{1+\frac{1}{x}} \\
&\overset{\text{QR}}{=\!=} \frac{\lim_{x\to\infty} 1}{\lim_{x\to\infty}\left(1+\frac{1}{x}\right)} \\
&\overset{\text{SR}}{=\!=} \frac{1}{1+0} = 1\,.
\end{aligned}$$

It worked! The conclusion is that $y=1$ is a horizontal asymptote.

Example 2.11.6: another horizontal asymptote

Evaluate the limit of this function at ∞:

$$\frac{x^2}{x^2+1}\,.$$

Substituting (mistake!) leads to

$$\frac{\infty}{\infty} \quad \boxed{\textbf{DEAD END}}$$

which is indeterminate. Let's divide by x as last time:

$$\begin{aligned}
\frac{x^2}{x^2+1} &= \frac{\frac{x^2}{x}}{\frac{(x^2+1)}{x}} \\
&= \frac{x}{x+\frac{1}{x}} \to \frac{\infty}{\infty} \quad \boxed{\textbf{DEAD END}}
\end{aligned}$$

It's still indeterminate! We haven't eliminated the infinities. Let's divide by x again,

$$\begin{aligned}
\frac{x}{x+\frac{1}{x}} &= \frac{\frac{x}{x}}{\frac{(x+\frac{1}{x})}{x}} \\
&= \frac{1}{1+\frac{1}{x^2}} \\
&\to \frac{1}{1+0} \\
&= 1\,.
\end{aligned}$$

Better idea: Divide by x^2 in the first place.

How do we evaluate the limit at ∞? The same way we did it for sequences in the last chapter. For a rational function, the *leading terms* of the two polynomials determine the long-term behavior of the numerator and denominator and, therefore, of the whole fraction.

Example 2.11.7: leading terms analysis

Consider:

$$\lim_{x\to\infty} \frac{\overbrace{3x^3}^{\text{long-term!}} \overbrace{-2x^2+x-8}^{\text{short-term...}}}{\underbrace{2x^2}_{\text{parabola!}} - \underbrace{17x+5}_{\text{where it is...}}} \to \frac{\infty}{\infty}\,.$$

This is what determines the long term:

$$\frac{3x^2}{2x^2} = \frac{3}{2}x \to \infty\,.$$

To make this visible, we divide numerator and denominator by x^2.

$$\begin{aligned}\lim_{x\to\infty} \frac{\frac{(3x^3-2x^2+x-8)}{x^2}}{\frac{(2x^2-17x+5)}{x^2}} &= \lim_{x\to\infty} \frac{3x-2+\frac{1}{x}-\frac{8}{x^2}}{2-\frac{17}{x}+\frac{5}{x^2}}\\ &= \frac{\infty-2+0-0}{2-0+0}\\ &= \infty\,.\end{aligned}$$

- *The lesson:* To get rid of the indeterminacy, get rid of one of the infinities.
- *The plan:* Divide both parts of the fraction by x to the degree of the denominator.

Example 2.11.8: rational

Analyze:

$$f(x) = \frac{x^3-x}{x^2-6x+5}\,.$$

At ∞, we have:

$$\begin{aligned}\lim_{x\to\infty} \frac{x^3-x}{x^2-6x+5} &= \lim_{x\to\infty} \frac{\frac{x^3-x}{x^2}}{\frac{x^2-6x+5}{x^2}}\\ &= \lim_{x\to\infty} \frac{x-\frac{1}{x}}{1-\frac{6}{x}+\frac{5}{x^2}}\\ &= \frac{\infty-0}{1-0-0}\\ &= \infty\,.\end{aligned}$$

It's infinite, yes, but which infinity? We need to find the *sign*. We just find the signs of the leading terms of the numerator and denominator.

$$\begin{array}{ll}\lim_{x\to+\infty} (x^3-x) = +\infty & \lim_{x\to-\infty} (x^3-x) = -\infty\\ \lim_{x\to+\infty} (x^2-6x+5) = +\infty & \lim_{x\to-\infty} (x^2-6x+5) = +\infty\end{array}$$

Conclusion: no horizontal asymptotes.

Now the vertical asymptotes. We look at

$$\frac{x^3-x}{x^2-6x+5}$$

and search for 0 of the denominator. Let's factor both:

$$\begin{aligned}x^3-x &= x(x^2-1) = x(x-1)(x+1)\\ x^2-6x+5 &= (x-1)(x-5)\end{aligned}$$

We use the quadratic formula:

$$\begin{aligned} x &= \frac{6 \pm \sqrt{6^2 - 4 \cdot 5}}{2} = \frac{6 \pm \sqrt{36 \cdot 20}}{2} \\ &= \frac{6 \pm 4}{2} = 5,\ 1\,. \end{aligned}$$

Now, we have:

$$f(x) = \frac{x(x-1)(x+1)}{(x-5)(x-1)} = \frac{x(x+1)}{x-5}$$

for $x \neq 1$ (that's the domain!). Next:

$$\lim_{x \to 1} f(x) = \frac{1(1+1)}{1-5} = -\frac{1}{2}\,,$$

because $\dfrac{x(x+1)}{x-5}$ is continuous at 1. So, $x = 1$ is *not* a vertical asymptote. Next:

$$\lim_{x \to 5^-} \frac{x(x+1)}{x-5} = \frac{30}{0} \stackrel{?}{=} \pm\infty\,.$$

Therefore, $x = 5$ is a vertical asymptote. But which infinity? We need to find the sign:

$$\frac{+ \cdot +}{-} = -$$

We conclude:

$$\lim_{x \to 5^-} \frac{x(x+1)}{x-5} = -\infty\,.$$

Similarly, we find:

$$\lim_{x \to 5^+} f(x) = +\infty\,.$$

Now some facts about limits of sequences can be restated for functions.

First, recall the lesson about the behavior of polynomials at ∞:

- ▶ Only the leading term matters.

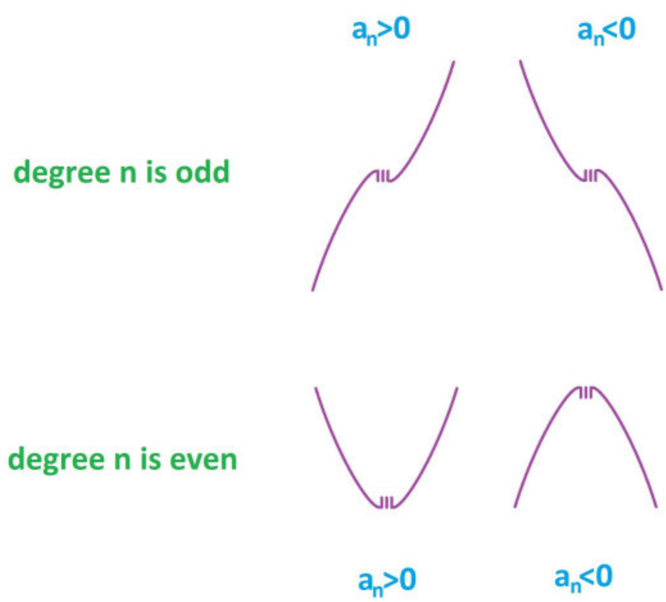
The complete statement is as follows:

Theorem 2.11.9: Limits of Polynomials at Infinity

Suppose we have a polynomial of degree p with the leading coefficient $a_p \neq 0$. Then its limit is as follows:

$$\lim_{x \to \pm\infty} (a_p x^p + a_{p-1} x^{p-1} + ... + a_1 x + a_0) = \begin{cases} \pm\infty & \text{if } a_p > 0 \\ \mp\infty & \text{if } a_p < 0 \end{cases}$$

Then, if we zoom out on the graph of a polynomial, we will see just four possible patterns



A more general result is about rational functions:

Theorem 2.11.10: Limits of Rational Functions at Infinity

Suppose we have a rational function represented as a quotient of two polynomials of degrees p and q, with the leading coefficients $a_p \neq 0,\ b_q \neq 0$. Then its limit is as follows:

$$\boxed{\lim_{x\to+\infty} \frac{a_p x^p + a_{p-1}x^{p-1} + ... + a_1 x + a_0}{b_q x^q + b_{q-1}x^{q-1} + ... + b_1 x + b_0} = \begin{cases} \infty & \text{if } p > q \\ \dfrac{a_p}{b_p} & \text{if } p = q \\ 0 & \text{if } p < q \end{cases}}$$

The last two cases produce horizontal asymptotes.

Exercise 2.11.11

State the above theorem for $x \to -\infty$.

The long-term behavior of a rational function is determined by the leading terms of its numerator and denominator:

$$\boxed{\lim_{x\to\infty} \frac{a_p x^p + a_{p-1}x^{p-1} + ... + a_1 x + a_0}{b_q x^q + b_{q-1}x x^{q-1} + ... + b_1 x + b_0} = \lim_{x\to\infty} \frac{a_p x^p}{b_p x^q} = \frac{a_p}{b_p} \lim_{x\to\infty} x^{p-q}}$$

The picture is complicated – in comparison to polynomials – by the fact that rational functions also have horizontal and vertical asymptotes.

Example 2.11.12: not rational

Sometimes we face indeterminate expressions with functions other than rational. Evaluate:

$$\lim_{x\to\infty} \left(\sqrt{x^2+1} - x\right).$$

Substituting (mistake!) gives us:

$$\infty - \infty \quad \boxed{\textbf{DEAD END}}$$

This is indeterminate. The trick is to multiply and divide by the "conjugate" of this expression:

$$\sqrt{x^2+1} + x\,.$$

This is how it works:

$$\begin{aligned}\lim_{x\to\infty}\frac{\left(\sqrt{x^2+1}-x\right)\left(\sqrt{x^2+1}+x\right)}{\sqrt{x^2+1}+x} &= \lim_{x\to\infty}\frac{(x^2+1)-x^2}{\sqrt{x^2+1}+x}\\ &= \lim_{x\to\infty}\frac{1}{\sqrt{x^2+1}+x}\\ &= \frac{1}{\infty}\\ &= 0\,.\end{aligned}$$

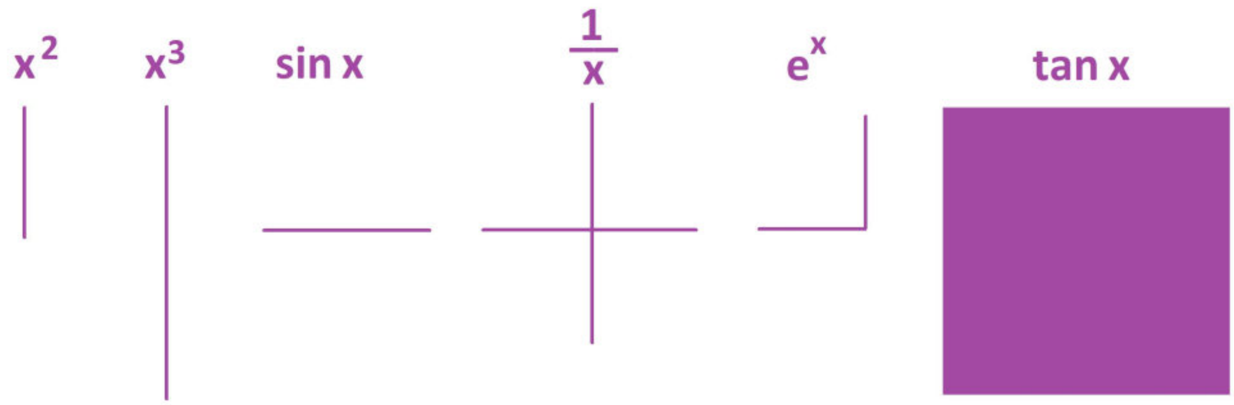
Exercise 2.11.13

Evaluate the limit: $\lim_{x\to\infty}\ln\left(\frac{x}{x+1}\right)$.

Exercise 2.11.14

Give an example of a function for each of the following patterns as $x\to\infty$: (a) $f(x)\to -1$, (b) $f(x)\to .33$, (b) $f(x)\to +\infty$, (c) $f(x)$ diverges but not to infinity.

In conclusion, this is what the graphs of some familiar functions look like at an extreme distance:



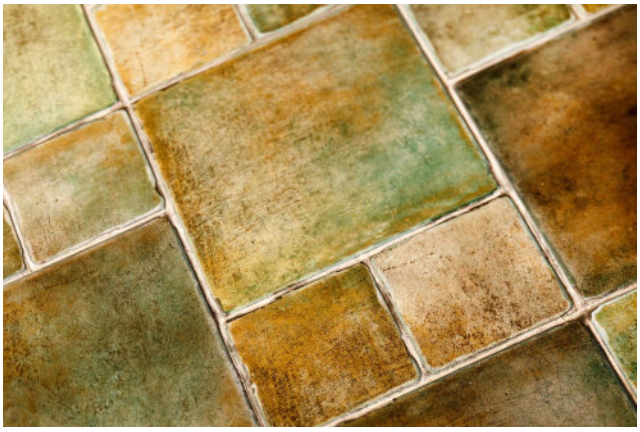
Exercise 2.11.15

Plot in this manner the graphs of the functions in this section.

2.12. Continuity and accuracy

The idea of continuity can be introduced and justified by considering the accuracy of a measurement.

Suppose we have a collection of square tiles of various sizes and we need to find the area A of each of them in order to know how many we need to cover the whole floor.



The answer is, of course, to measure the side, x, of each tile and then compute:

$$A = x^2 \,.$$

For example, we have:

$$x = 10 \implies A = 100 \,.$$

But measurements are never fully accurate; for example, we might have:

$$x = 10 \pm .3 \,.$$

In other words, x lies within a certain interval:

$$x \text{ in } (10 - .3, 10 + .3) = (9.7, 10.3) \,.$$

As a result, the computed value of the area of the tile – what we care about – will also have some error! Indeed, the area won't be just $A = 100$ but

$$A = (10 \pm .3)^2 \,.$$

In other words, A lies within a certain interval:

$$A \text{ is in } \left((10 - .3)^2, (10 + .3)^2\right) = (9.7^2, 10.3^2) = (94.09, 106.09) \,.$$

This is informative, but our investigation goes in the opposite direction.

Suppose next that we can always improve *the accuracy of the measurement* of the side of the tile x – as much as we like. The question is, can we also improve *the accuracy of the computed value* of A. Can we achieve this accuracy to anybody's satisfaction, even if this standard of accuracy might change?

Suppose again that $x = 10$. If the desired accuracy of A is ± 3, we *require*:

$$A \text{ is in } (100 - 3, 100 + 3) = (97, 103) \,.$$

But this interval doesn't contain the interval we found to guarantee to contain A: $(94.09, 106.09)$! Therefore, we haven't achieved the desired accuracy of A with the given accuracy of measurement x, which is $\pm .3$.

If the desired accuracy of A is ± 7, we have:

$$A \text{ is in } (94.09, 106.09) \subset (100 - 7, 100 + 7) = (93, 107) \,.$$

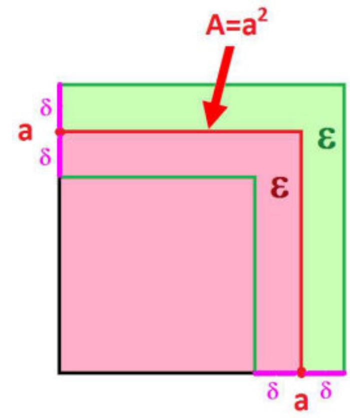
It is satisfied!

We can easily show that the error $\pm .1$ for x provides the original, smaller threshold for A:

$$A \text{ is in } \left((10 - .1)^2, (10 + .1)^2\right) = (9.9^2, 10.1^2) = (98.01, 102.01) \subset (100 - 3, 100 + 3) = (97, 103) \,.$$

Let's rephrase this problem in order to solve it for all possible values of the desired accuracy of A.

Let's assume that the measurement of the side is a and, therefore, the assumed area A is a^2. Now suppose we want the deviation of A from the true (unknown) area to be some small value $\varepsilon > 0$ or better. What threshold δ for the deviation of a from x do we need to be able to guarantee that? We want to ensure that A is within ε from a^2 by making sure that x is within δ from a:



We know this:

$$A \text{ is in } \left((a-\delta)^2, (a+\delta)^2\right).$$

We need to demonstrate that:

$$A \text{ is in } (a^2-\varepsilon, a^2+\varepsilon).$$

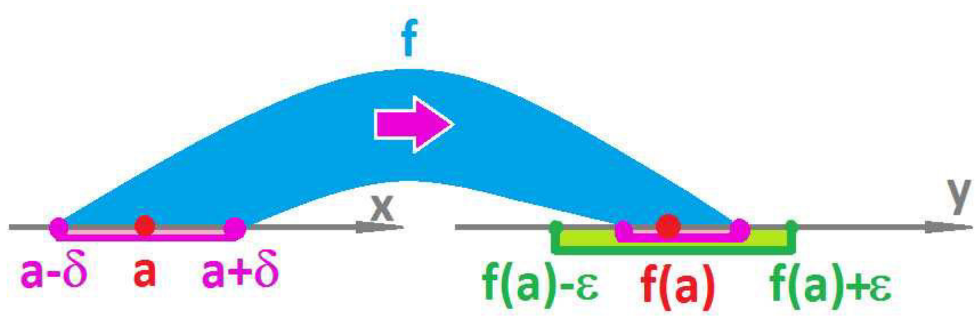
The latter will follow from the former, provided:

$$\left((a-\delta)^2, (a+\delta)^2\right) \subset (a^2-\varepsilon, a^2+\varepsilon).$$

More generally, we have the following inclusion for any increasing function f:

$$\left(f(a-\delta), f(a+\delta)\right) \subset (f(a)-\varepsilon, f(a)+\varepsilon).$$

This inclusion is illustrated below:



We rephrase the last inclusion of intervals:

$$x \text{ within } \delta \text{ from a } \implies f(x) \text{ within } \varepsilon \text{ from f(a) }.$$

The definition suggested by the above discussion is as follows:

Definition 2.12.1: ε-δ definition of continuity

A function f is called *continuous* at $x=a$ if for any $\varepsilon>0$ there is $\delta>0$ such that

$$|x-a|<\delta \implies |f(x)-f(a)|<\varepsilon.$$

If we can demonstrate that this new definition is equivalent to the one presented in this chapter, we can answer our question in the affirmative:

► Yes, we *can* always improve the accuracy of the computed value of $A=x^2$ – to anybody's satisfaction – by improving the accuracy of the measurement of x.

Exercise 2.12.2

Prove that $f(x)=x^2$ is continuous at $x=0,\ x=1,\ x=a$.

Exercise 2.12.3

Carry out this kind of analysis for: a thermometer put in a cup of coffee to find its temperature. Assume that the thermometer gives perfect readings. Hint: It'll take time for it to warm up.

Exercise 2.12.4

What is the relation between ε and δ when f is *linear*?

To further illustrate the idea, let's consider a different situation. Suppose we don't care about the area anymore; we just want to fit these tiles into a strip 10 inches wide. We will be testing the tiles and decide ahead of time whether it will fit: If it fits, it is used; otherwise, it is discarded.

Then, we still get a measurement a of the side of the tile but our real interest is whether a is less or more than 10. Just as in the previous example, we don't know the actual length x exactly; it's always within some limits: 10.0 ± 0.5 or $a \pm \delta$. Here δ is the accuracy of measurement of a. The algebra is much simpler than before. For example, if the length is measured as 11, we need the accuracy $\delta = 1$ or better to make the determination. It's the same for the length 9.

But what if the measurement is exactly $a = 10$? Even if we can improve the accuracy, i.e., δ, as long as $\delta > 0$, we can't know whether x is larger or smaller than 10.

Let's define a function that makes this decision:

$$f(x) = \begin{cases} 1 & \text{(pass)} \qquad \text{if } x \le 10\,, \\ 0 & \text{(fail)} \qquad \text{if } x > 10\,. \end{cases}$$

Suppose we need the accuracy of $y = f(x)$ to be $\varepsilon = 0.5$. Can we achieve this by decreasing δ? In other words, can we find δ such that

$$|x - 10| < \delta \implies |f(x) - 1| < \varepsilon\,?$$

Of course not:

$$x > 10 \implies |f(x) - 1| = |0 - 1| = 1\,.$$

Thus, the answer to our question is:

► No, we *cannot* always improve the accuracy of the computed value of $f(x)$ – to anybody's satisfaction – by improving the accuracy of the measurement of x.

The reason to be quoted is that f is discontinuous at $x = 10$.

Exercise 2.12.5

Carry out this kind of analysis for: the total test score vs. the corresponding letter grade. What if we introduce A-, B+, etc.?

We next pursue this idea of continuity of the dependence of y on x:

► We can ensure the desired accuracy of y by increasing the accuracy of x.

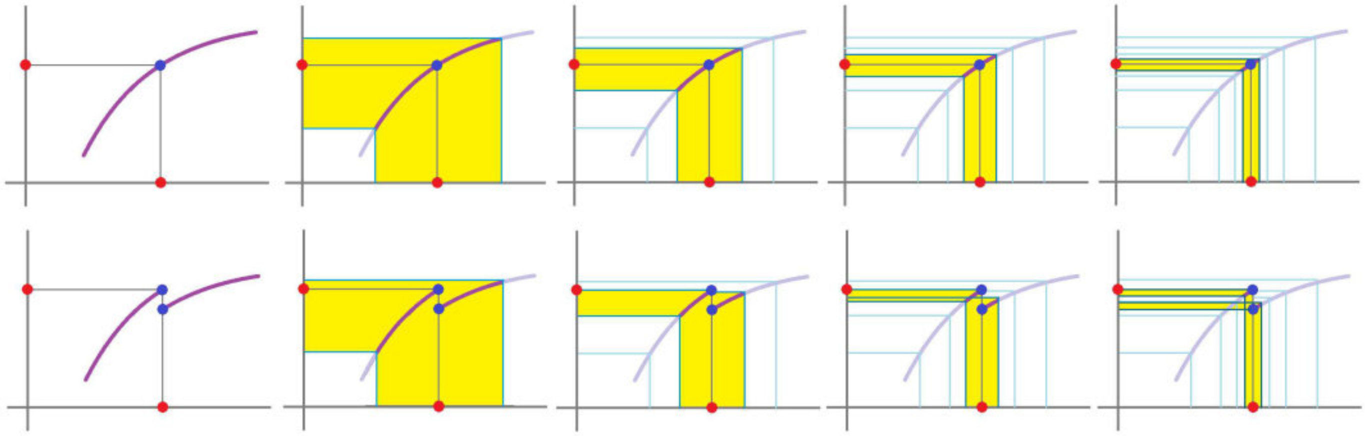
We discuss limits first.

2.13. The ε-δ definition of limit

We can define limits and continuity of functions without invoking limits of sequences. Let's start over.

Example 2.13.1: images of intervals

Let's compare what happens to open *intervals* that contain our point of interest under a function. First, suppose the function is continuous (top). As the interval in the x-axis shrinks, so does the corresponding interval in the y-axis:



Second, consider this discontinuous function (bottom). As the interval in the x-axis shrinks, we notice two things about the corresponding set in the y-axis:

- It eventually ceases to be an interval. A tear!
- It also stops shrinking. A gap!

Below, we rewrite what we want to say about the meaning of the limits in progressively more and more precise terms.

x	$y = f(x)$
As $x \to a$,	we have $y \to L$.
As x approaches a,	y approaches L.
As the distance from x to a approaches 0,	the distance from y to L approaches 0.
As $\lvert x - a \rvert \to 0$,	we have $\lvert y - L \rvert \to 0$.
By making $\lvert x - a \rvert$ as smaller and smaller,	we make $\lvert y - L \rvert$ as small as needed.
By making $\lvert x - a \rvert$ less than some $\delta > 0$,	we make $\lvert y - L \rvert$ smaller than any given $\varepsilon > 0$.

Now we have it:

Definition 2.13.2: ε-δ definition of limit

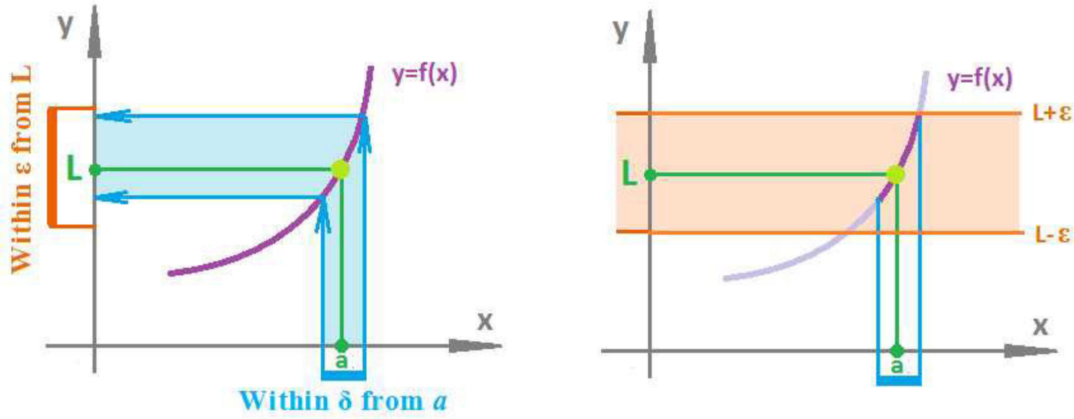
The *limit of a function* f at $x = a$ is a number L, if it exists, such that for any $\varepsilon > 0$ there is such a $\delta > 0$ that

$$0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon\,.$$

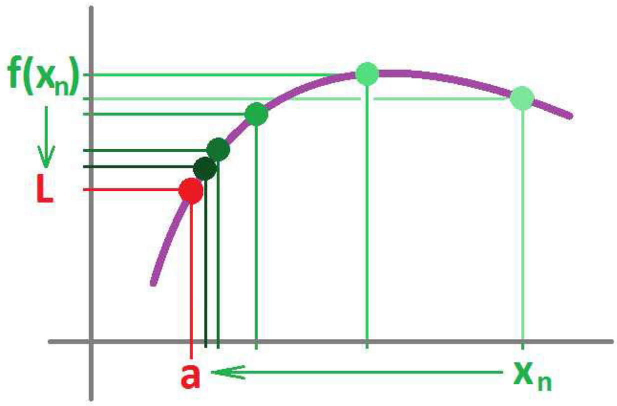
This is the geometric meaning of the definition:

▶ If x is within δ from a, then $f(x)$ is supposed to be within ε from L.

In other words, this part of the graph fits within the ε-band around the line $y = L$:



This is the "ε-δ definition" of limit of a function. It matches the definition of limit of a sequence, which can now be called the "ε-N definition".



This is our conclusion:

Theorem 2.13.3: Equivalence of Definitions of Limit

The two definitions of the limit of a function are equivalent.

Proof.

[ε-$\delta \Rightarrow \varepsilon$-$N$] Suppose $x_n \to a$. Suppose now that $\varepsilon > 0$ is given. As the definition above is satisfied, there is a $\delta > 0$ such that

$$0 < |x - a| < \delta \implies |f(x) - L| < \varepsilon\,.$$

On the other hand, we have:

$$|x_n - a| < \delta \text{ for all } n > N$$

for some N. Therefore, for all $n > N$, we have:

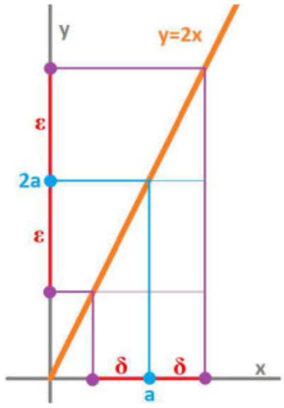
$$|f(x_n) - L| < \varepsilon\,.$$

This means that $f(x_n) \to L$.

Exercise 2.13.4

Prove the other half: [ε-$N \Rightarrow \varepsilon$-$\delta$].

As an exercise, let's compute the limit of an arbitrary linear function. The relation between δ and ε is predicable:



This is what we guess from the picture:

Theorem 2.13.5: Limit of Linear Function

For a linear function $f(x) = mx+b$ with $m \neq 0$, the definition of limit is satisfied for

$$\delta = \varepsilon/|m|\,.$$

Proof.

$$\begin{aligned}
|x-a| < \delta &\iff |x-a| < \varepsilon/|m| \\
&\iff |m|\cdot|x-a| < \varepsilon \\
&\iff |m\cdot(x-a)| < \varepsilon \\
&\iff |mx-ma| < \varepsilon \\
&\iff |(mx+b)-(ma+b)| < \varepsilon \\
&\iff |f(x)-f(a)| < \varepsilon\,.
\end{aligned}$$

Example 2.13.6: using the definition – linear

Prove the limit by using the definition:

$$\lim_{x\to -1.5} \frac{9-4x^2}{3+2x} = 6\,.$$

Let's match what we want to prove with the definition:

$$\lim_{x\to \underbrace{-1.5}_{a}} \underbrace{\frac{9-4x^2}{3+2x}}_{f(x)} = \underbrace{6}_{L}\,.$$

We replace $f(x)$ with a simpler expression, equal for all $x \neq -1.5$:

$$\begin{aligned}
\frac{9-4x^2}{3+2x} &= \frac{3^2-(2x)^2}{3+2x} \\
&= \frac{(3+2x)(3-2x)}{3+2x} \\
&= 3-2x
\end{aligned}$$

for all x for which $3+2x \neq 0$, or $x \neq 1.5$.

Now, this expression is linear! Then the dependence of δ on ε is simple, as we know:

$$\delta = \frac{\varepsilon}{2}\,.$$

We need to demonstrate the following:

$$0 < |x - (-1.5)| < \frac{\varepsilon}{2} \overset{?}{\Longrightarrow} |3 - 2x - 6| < \varepsilon$$

We rewrite it:

$$\begin{aligned}
0 < |x + 1.5| < \frac{\varepsilon}{2} &\overset{?}{\Longrightarrow} \quad |-3 - 2x| < \varepsilon \\
0 < |x + 1.5| < \frac{\varepsilon}{2} &\overset{?}{\Longrightarrow} \quad |2x + 3| < \varepsilon \\
0 < |x + 1.5| < \frac{\varepsilon}{2} &\overset{?}{\Longrightarrow} \quad |x + 1.5| < \frac{\varepsilon}{2}
\end{aligned}$$

It works! The definition is satisfied.

Example 2.13.7: using the definition – non-linear

Let's consider first $f(x) = x^2$ at $a = 0$. We expect the limit to be $L = 0$. But what is the value of δ in terms of ε?

$$\begin{aligned}
|f(x) - L| < \varepsilon \quad &\Longleftrightarrow \quad |x^2| < \varepsilon \\
&\Longleftrightarrow \quad |x|^2 < \varepsilon \\
&\Longleftrightarrow \quad |x| < \sqrt{\varepsilon}\,.
\end{aligned}$$

Thus, we have discovered that $\delta = \sqrt{\varepsilon}$ works in the definition.

Example 2.13.8: using the definition – non-linear

However, this is how changing the location from $a = 0$ to $a = 1$ complicates things:

$$\begin{aligned}
|f(x) - L| < \varepsilon \quad &\Longleftrightarrow \quad |x^2 - 1| < \varepsilon \\
&\Longleftrightarrow \quad |(x - 1)(x + 1)| < \varepsilon \\
&\Longleftrightarrow \quad |x - 1| < \frac{\varepsilon}{|x + 1|} \\
&\Longleftarrow \quad |x - 1| < \delta < \frac{\varepsilon}{|x + 1|}\,.
\end{aligned}$$

To find such a δ, we have to proceed to the estimating of the value of $|x + 1|$, just as we did before. The reason for this complication is explained by the fact that the graph of $y = x^2$ curves differently ($f'' \neq 0$) on the left and on the right of $a = 1$.

Example 2.13.9: DNE

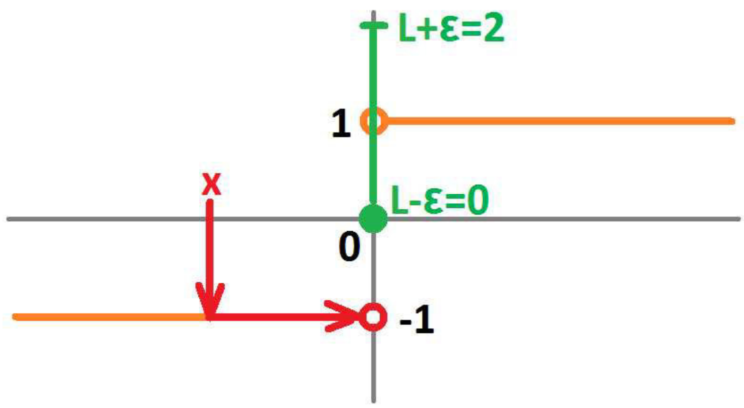
Let's consider the sign function around 0:

$$f(x) = \operatorname{sign}(x) = \begin{cases} -1 & \text{if } x < 0, \\ 0 & \text{if } x = 0, \\ 1 & \text{if } x > 0\,. \end{cases}$$

Our guess is that $\lim_{x \to 0} f(x)$ does not exist but for now, we will just show that $L = 1$ is *not the limit.* We need to show that there is no δ satisfying this condition for some $\varepsilon > 0$. Pick $\varepsilon = 1$. Is there $\delta > 0$ such that:

$$0 < |x - \overbrace{0}^{a}| < \delta \quad \Longrightarrow \quad \underbrace{|\overbrace{\operatorname{sign}(x)}^{f(x)} - \overbrace{1}^{L}| < \varepsilon = 1}_{f(x) \text{ is within } (0,2)}\,.$$

What x would violate this?



Any $x < 0$ since we have then $f(x) = -1$. The definition fails, so $L = 1$ is not the limit.

Exercise 2.13.10

Prove that $L = 0$ isn't the limit either.

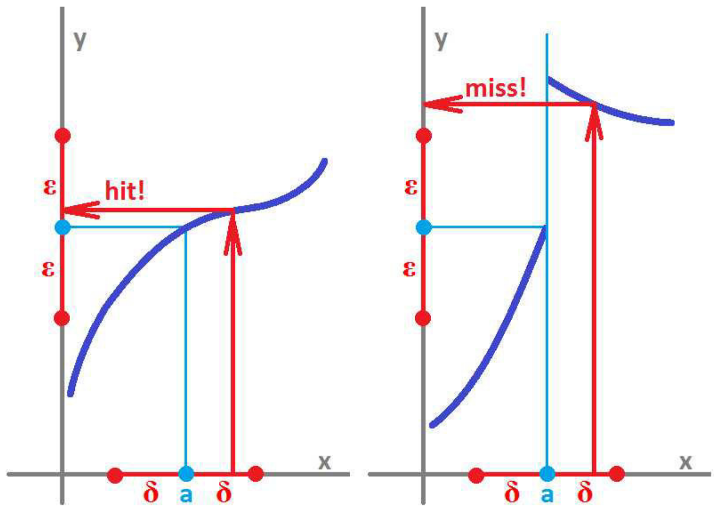
In order to learn how to *disprove* the existence of a limit, let's state the negation of the definition:

Theorem 2.13.11: Not Limit

A number L is not the limit of function f at $x = a$ if there is an $\varepsilon > 0$ such that for any $\delta > 0$ and some x we have:

$$0 < |x - a| < \delta \ \text{ and } \ |f(x) - L| \geq \varepsilon \,.$$

The existence of such an ε indicates the presence of a gap in the graph:



Example 2.13.12: sign function

Let's prove the non-existence of limit at $a = 0$ for the sign function $f(x) = \operatorname{sign}(x)$. Suppose $L \neq 0$ is *any* number. Let's pick $\varepsilon = |L|/2$. Now suppose $\delta > 0$ is arbitrary and suppose for some $x \neq 0$ we have:

$$0 < |x| < \delta \quad \text{and} \quad |f(x) - L| < \varepsilon = |L|/2 \,.$$

Then, $f(x)$ is of the same sign as L. But what does this say about $-x$? We also have:

$$0 < |-x| < \delta \text{ and, therefore, } |f(-x) - L| < \varepsilon = |L|/2 \,.$$

Then, $f(-x)$ is also of the same sign as L. This contradicts the fact that $f(-x) = -f(x)$.

Let's prove the *Composition Rule* for this new definition of limit.

We have the composition of functions f and g. These are the three instances of the definition we face; we need to prove part 3 from parts 1 and 2:

1. $y = g(x) \to L$ as $x \to a$, i.e., for any $\cancel{\varepsilon}\,\gamma > 0$ there is such a $\delta > 0$ that
$$0 < |x - a| < \delta \implies |g(x) - L| < \cancel{\varepsilon}\,\gamma\,.$$

2. $z = f(y) \to M$ as $y \to L$, i.e., for any $\varepsilon > 0$ there is such a $\cancel{\delta}\,\gamma > 0$ that
$$0 < |y - L| < \cancel{\delta}\,\gamma \implies |f(y) - M| < \varepsilon\,.$$

3. $z = f(g(x)) \to M$ as $x \to a$, i.e., for any $\varepsilon > 0$ there is such a $\delta > 0$ that
$$0 < |x - a| < \delta \implies |f(g(x)) - M| < \varepsilon\,.$$

Above we reconcile the parameters, ε and δ, that appear in each, by matching the three variables: A small deviation of x (how small: within δ) causes a small deviation of y (call it γ), which in turn causes a small deviation of z (that's ε).

Suppose now $\varepsilon > 0$ is given. Then the γ found, from this ε, in part 2 is fed into part 1, producing a δ. Combining these two statements together, we have:

$$\begin{array}{llll} 0 < |x - a| < \delta & \implies |g(x) - L| < \gamma\,. & \text{Let } y = g(x) & \\ & \implies |y - L| < \gamma & \implies |f(y) - M| < \varepsilon\,. \end{array}$$

Part 3 is proven! We finally state the result:

Theorem 2.13.13: Limit under Composition

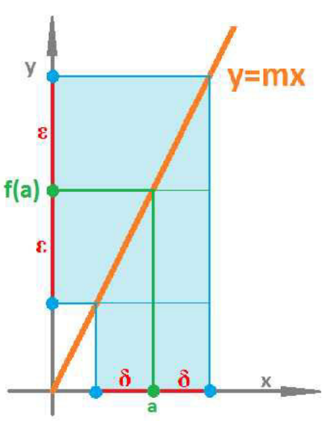
Suppose
1. $y = g(x) \to L$ *as* $x \to a$, *and*
2. $z = f(y) \to M$ *as* $y \to L$.

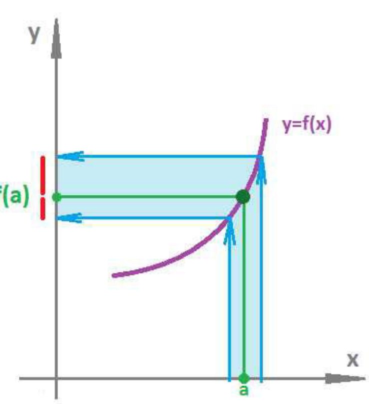
Then
$$z = (f \circ g)(x) \to M \quad \text{as } x \to a\,.$$

Exercise 2.13.14

Prove that if f and g are continuous at $x = a$, then so are $\max\{f, g\}$ and $\min\{f, g\}$.

We have required that if x is within δ from a, then $f(x)$ is within ε from L. We are referring here to intervals – one containing a and another L – but not just any; these intervals are *centered* at these points. This symmetry requirement isn't necessary:





Even when we start with a symmetric interval around a, the symmetry around L occurs only when the function is *linear*:

Theorem 2.13.15: Alternative Definition of Limit

The limit of a function f at $x = a$ is L if and only if for any $\varepsilon_1,\ \varepsilon_2 > 0$ there are such $\delta_2,\ \delta_2 > 0$ that

$$-\delta_1 < x - a < \delta_2 \implies -\varepsilon_1 < f(x) - L < \varepsilon_2\,.$$

The theorems supply an alternative definition of limit, and of continuity.

Exercise 2.13.16

(a) Prove the above theorem. (b) Use the above theorem to prove the *Composition Theorem* above. (c) Use the above theorem to prove that the inverse of a continuous function is continuous.

Chapter 3: The derivative

Contents

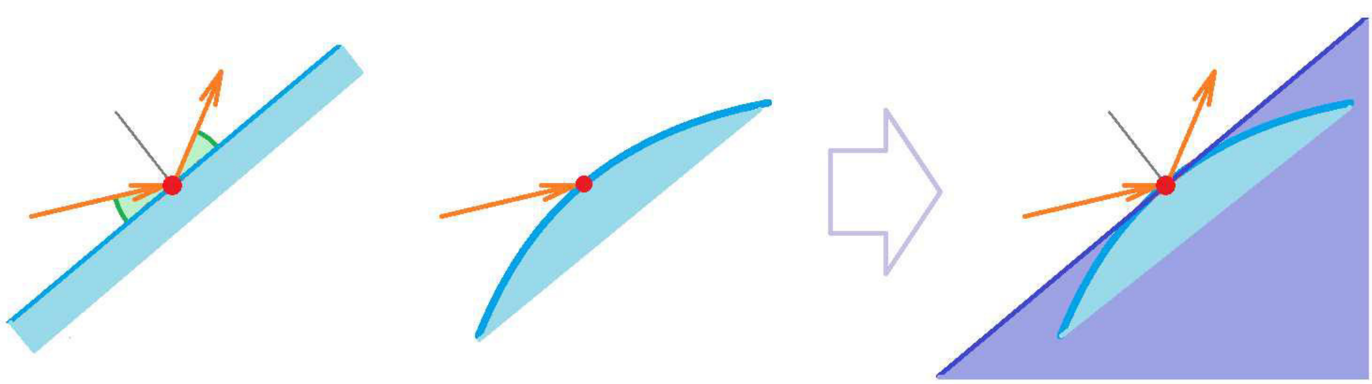
3.1. The Tangent Problem

There are two main ways to enter calculus:

- through the study of motion, and
- through the study of curved shapes ("geometry").

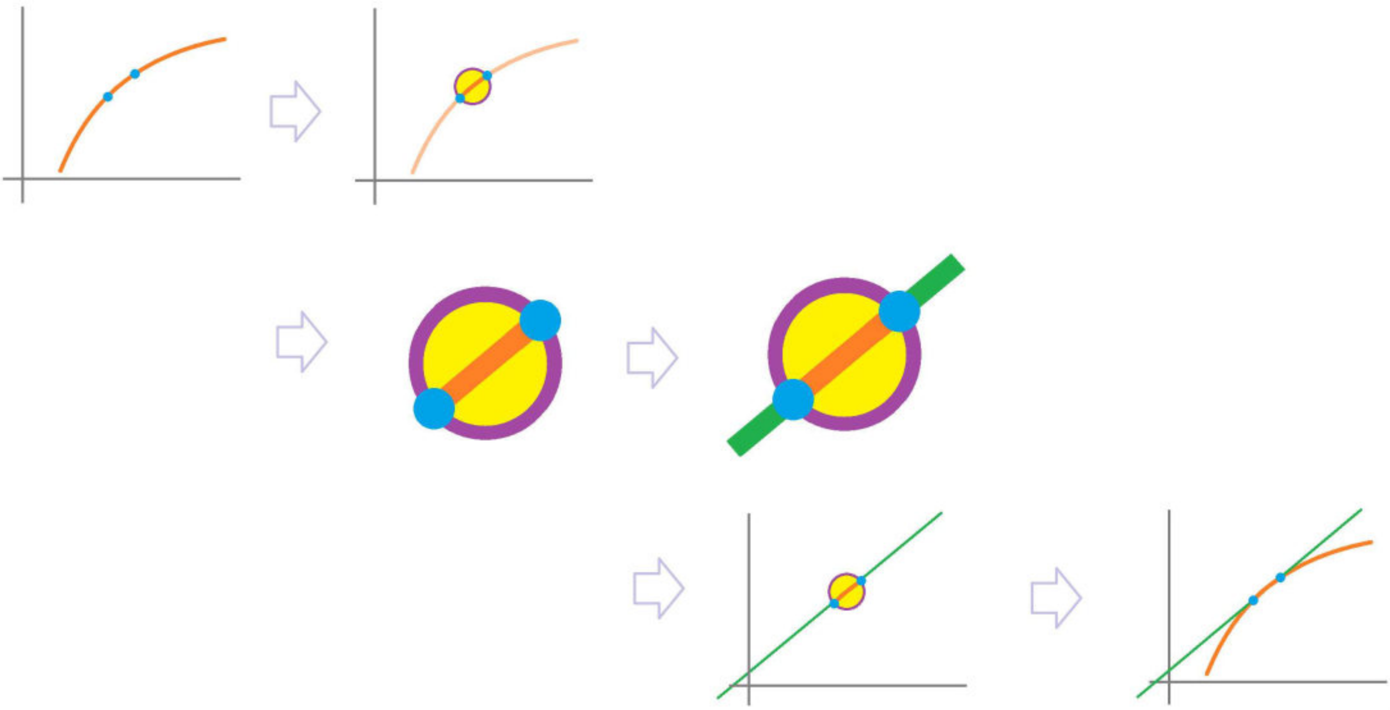
We started with the former in Chapter 1 and will continue on shortly. In this section, we consider the latter.

In what direction will light bounce off a mirror? We know the answer when the mirror is straight (left), but what about a curved mirror (middle)?



The ray is extremely narrow and there is just a single point of contact. Therefore, the light bounces as if off a *straight* mirror at the point of contact (right). To see that, we zoom in on the point.

What do we see exactly? The first possibility is that we might see two points connected by a straight edge:



We assume that the light will hit this edge and this is the one we use to find its path after the contact. This line cuts across the curve and is called a *secant* (the Greek for "cut") line.

How do we find these lines? Since a point is supplied for each, all we need is the angle. In the presence of the Cartesian system, we use the *slopes*, and then the point-slope form of the line, to find those lines (seen in Volume 1, Chapter 1PC-2).

Let's review the most basic concept:

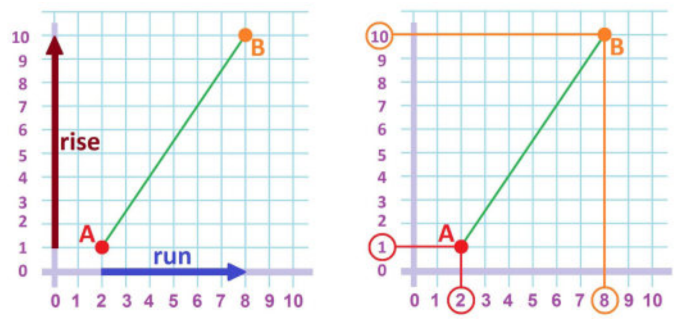
Definition 3.1.1: slope

Suppose we have two points in a specified order, $A = (x_0, y_0)$ then $B = (x_1, y_1)$, on the xy-plane, then *the slope of the line* from A to B is defined to be

$$\boxed{\text{slope } = m = \frac{\text{rise}}{\text{run}} = \frac{\text{change of } y}{\text{change of } x} = \frac{\text{change from } y_0 \text{ to } y_1}{\text{change from } x_0 \text{ to } x_1} = \frac{y_1 - y_0}{x_1 - x_0}}$$

Example 3.1.2: computing slopes

The geometric meaning of the numerator and denominator is seen below:



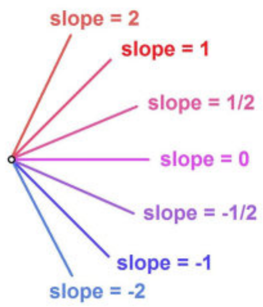
Here we have:
- run $= 8 - 2 = 6$, and
- rise $= 10 - 1 = 9$, therefore,
- slope $= \dfrac{9}{6} = \dfrac{3}{2} = 1.5$.

Exercise 3.1.3

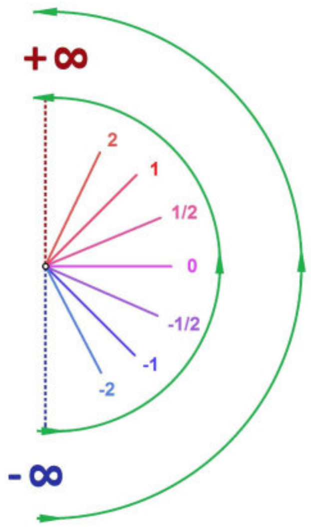
Express the slope in trigonometric terms.

What does the slope tell us about the line?

Below we arrange all linear functions according to their slopes (with the same y-intercept):

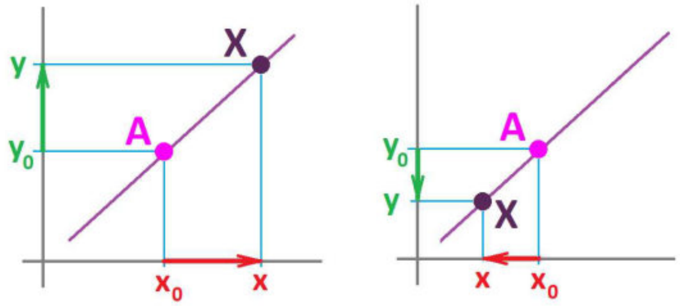


It's as if increasing the slope *rotates* the line counterclockwise:



It can't make the line vertical though.

Suppose we have a specified point $A = (x_0, y_0)$ on our line. Let's consider an arbitrary point $X = (x, y)$ on the line:



The run is $x - x_0$ and the rise is $y - y_0$ (left or right). Therefore, the slope is

$$m = \frac{y - y_0}{x - x_0}.$$

In this formula, x cannot be equal to x_0. It cannot, therefore, be used as a formula for the line! Let's change our view on the slope: from

$$\text{slope} \ = \frac{\text{rise}}{\text{run}}$$

to

$$\text{rise} \ = \ \text{slope} \ \cdot \ \text{run}\,.$$

We have a new way to represent a line:

Theorem 3.1.4: Point-Slope Form of Line

A line with slope m passing through point (x_0, y_0) is given by the following linear relation:

$$\boxed{y - y_0 = m \cdot (x - x_0)}$$

Below is the breakdown of the formula:

Point-slope form of line

$$\begin{array}{ccccc} \text{rise} & = & \text{slope} & \cdot & \text{run} \\ (y - y_0) & = & m & \cdot & (x - x_0) \end{array}$$

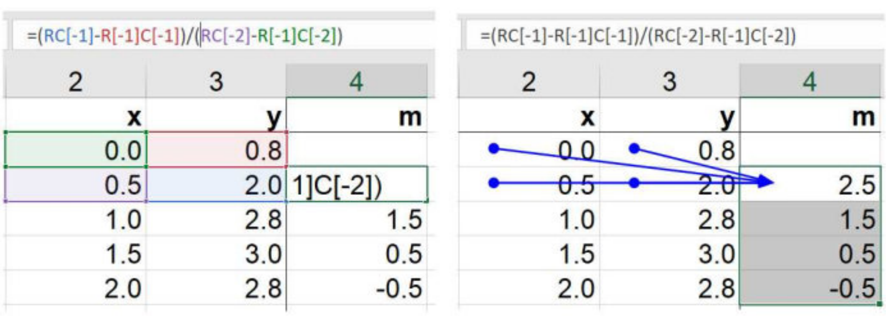
$$\begin{array}{ccccccccc} \text{point } X & & & & & & \text{point } X & & \\ \downarrow & & & & & & \downarrow & & \\ y & - & y_0 & = & m & \cdot & (x & - & x_0) \\ & & \uparrow & & & & & & \uparrow \\ & & \text{point } A & & & & & & \text{point } A \end{array}$$

Example 3.1.5: finding slopes

Suppose we have two columns of numbers x, y in a spreadsheet. This data is visualized as a curve given by its points on the xy-plane. Then we can carry out a computation of the slope for each consecutive pair of points with the following formula:

```
=(RC[-1]-R[-1]C[-1])/(RC[-2]-R[-1]C[-2])
```

The results are placed in the third column:



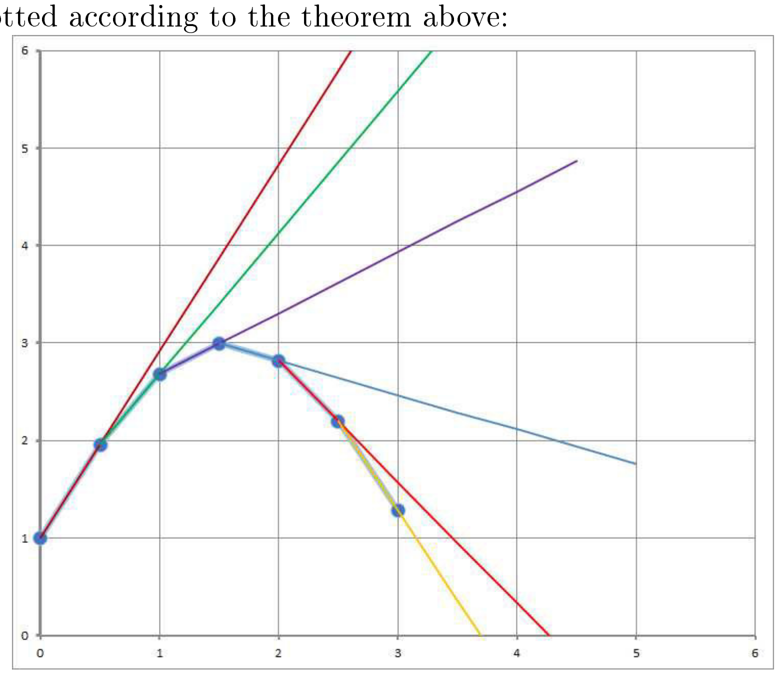
=(RC[-1]-R[-1]C[-1])/(RC[-2]-R[-1]C[-2])

2	3	4
x	y	m
0.0	0.8	
0.5	2.0	1]C[-2])
1.0	2.8	1.5
1.5	3.0	0.5
2.0	2.8	-0.5

=(RC[-1]-R[-1]C[-1])/(RC[-2]-R[-1]C[-2])

2	3	4
x	y	m
0.0	0.8	
0.5	2.0	2.5
1.0	2.8	1.5
1.5	3.0	0.5
2.0	2.8	-0.5

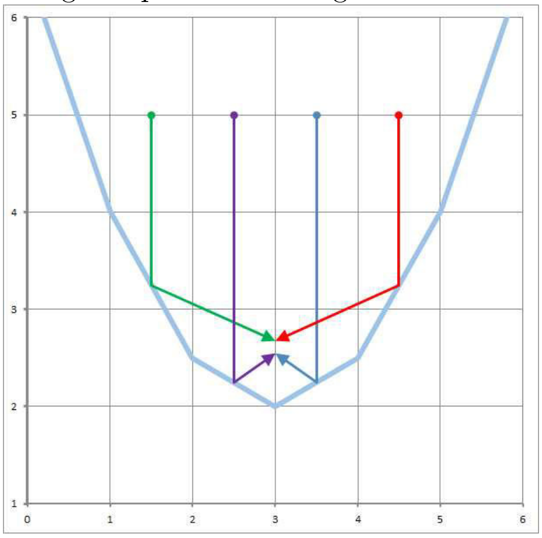
The lines are then plotted according to the theorem above:



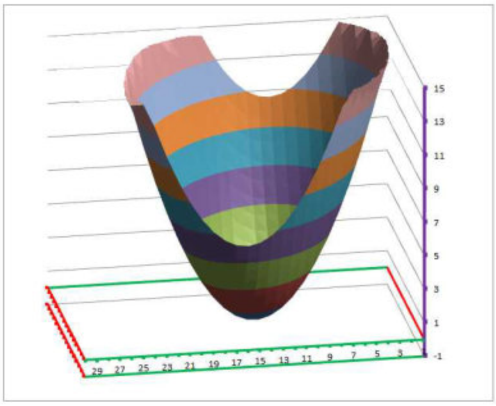
Example 3.1.6: parabola

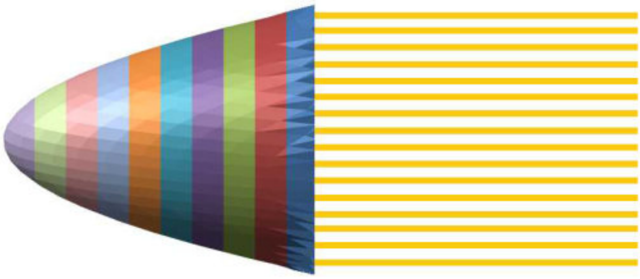
The history of the ancient Greece tells about the famous mathematician Archimedes who arranged the shields of the solders along this curve in order to set on fire the ships of the Romans that besieged his home city of Syracuse. He used the fact that the rays of light bouncing from a parabolic mirror will all meet at one point called the *focus* of the parabola.

We can confirm that idea by tracing the paths of the light beams as they bounce off the secant lines:

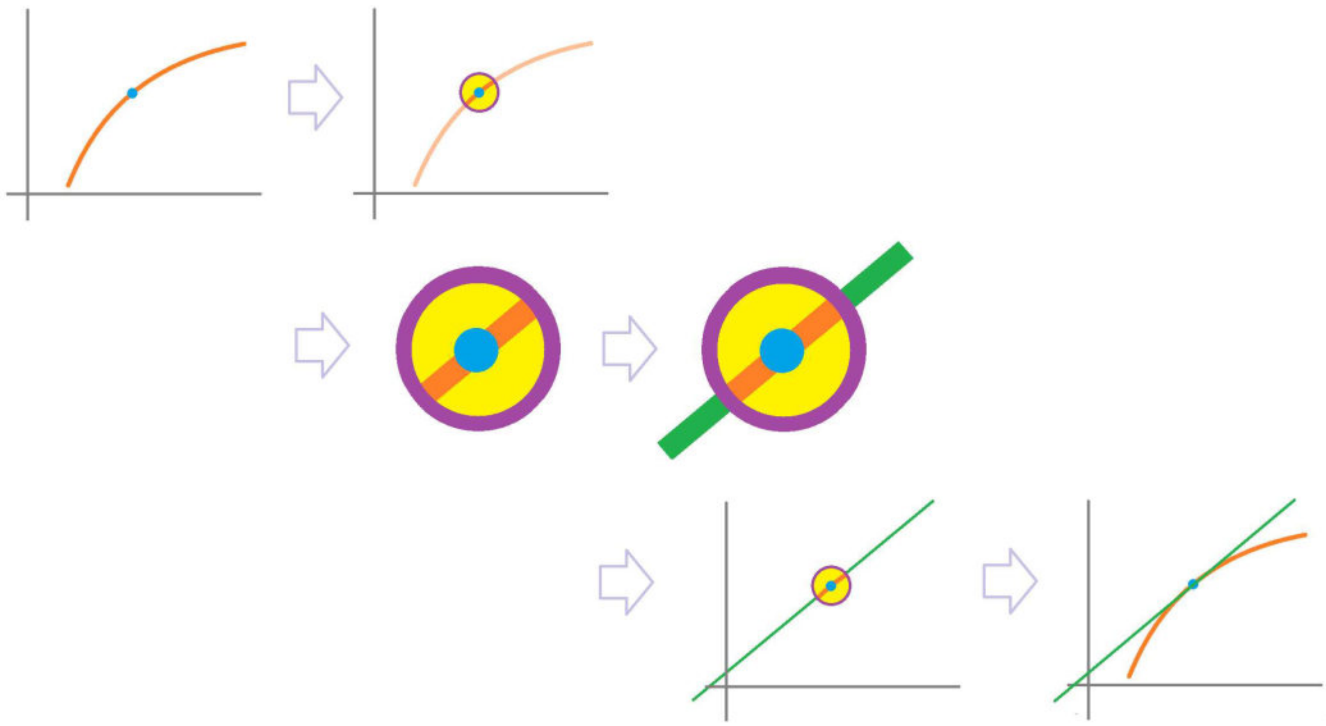


Conversely, a source of light placed at the focus of a parabola will create a beam of parallel rays of light; the fact is used to design cars' headlights:





Another hypothetical possibility is that we might see no corners even after zooming in multiple times: we deal with idealized, mathematical curves. In this case, zooming in produces a virtually straight line:

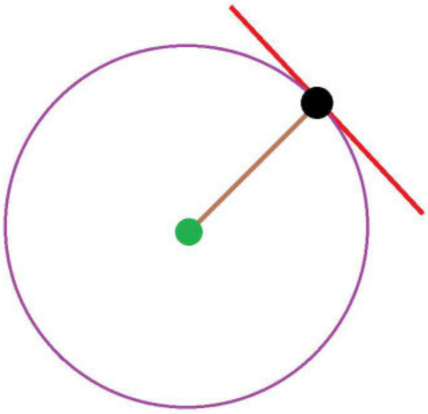


It is this line that the light will hit. Therefore, its path following the contact can be derived if this line is found. This line *touches* the curve and is called a *tangent* (the Greek for "touch") line.

How do we find these "touching" lines? It is much more complex.

Example 3.1.7: circle

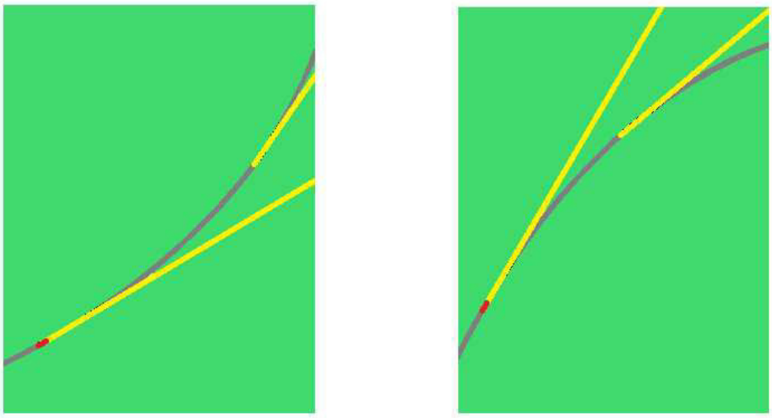
The problem of finding such a line has an easy solution for some specific curves, such as a circle:



Once they realized that the radius and such a line form a 90-degree angle, the problem was solved by the ancient Greeks with just ruler and compass. Indeed, one just plot another circle with diameter served by the segment between the point and the center of the original circle.

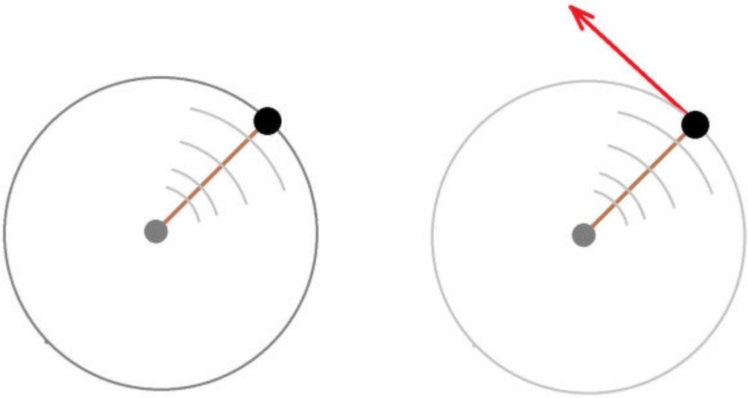
Example 3.1.8: headlights

Where do the lights of a car traveling on a curvy road point?



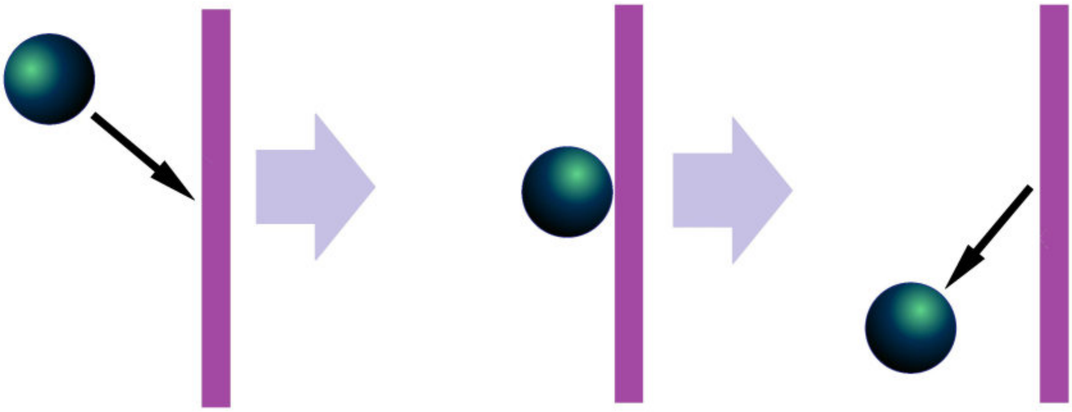
Example 3.1.9: sling

In what direction would a rock released from a sling go (view from above)?

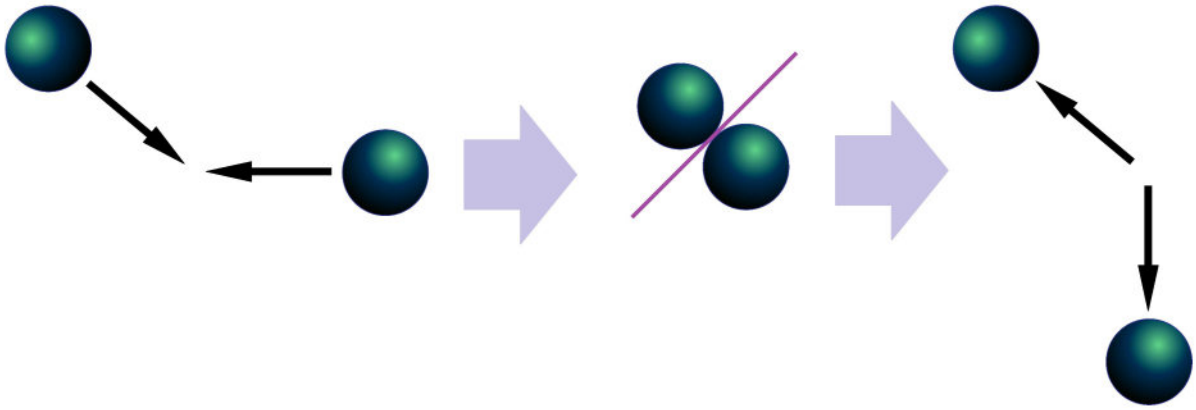


Answering the above question will also answer the questions below.

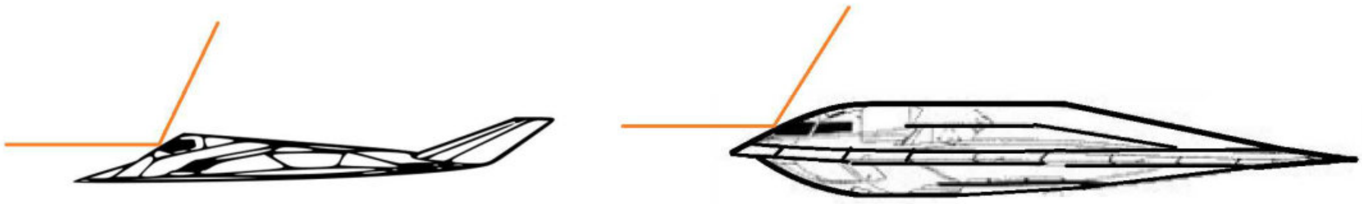
How would a billiard ball bounce off a wall? Same as light:



How would two balls bounce off each other? As if they both hit a wall:

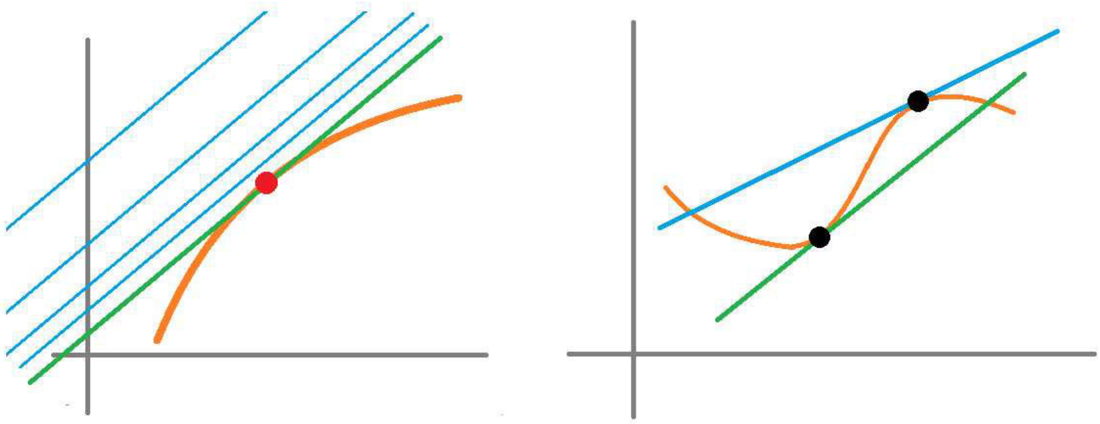


Next, in which direction will a radar signal bounce off the surface of a plane? If you want to know ahead of time, design a plane with no curves:

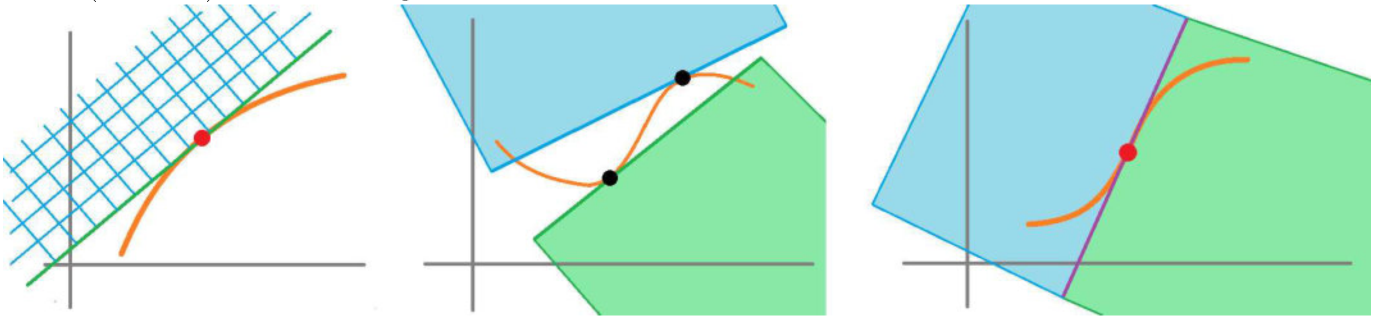


Example 3.1.10: using a ruler

Practically, how do we find the tangent line of a curve drawn on a piece of paper? As we don't have the luxury of zooming in on a digital image, we draw straight lines closer and closer to the graph so that the last one touches the graph at *that* point.



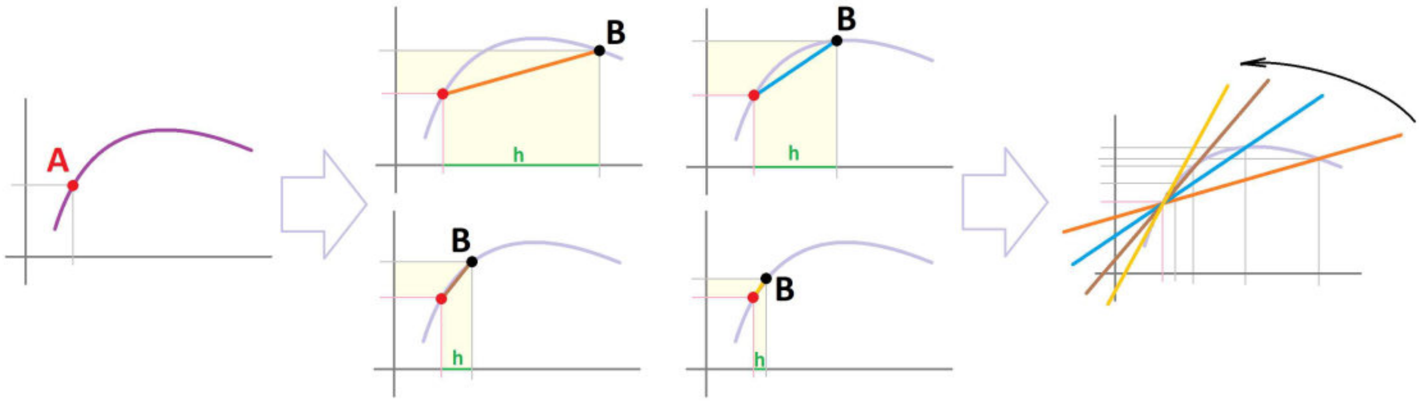
The rule of thumb is that one should expect – when zoomed out – only one point in common between the tangent line and the part of the graph of f close to the point A. Another way to know that you did it right is to see the tangent line as an edge of a piece of paper; then this piece has to cover none of the (relevant) parts of the graph:



There are exceptions; the last image gives you an *inflection point* with the tangent line cutting the graph in half.

The analytical method of solving this *Tangent Problem* is one of the two main motivations (the other one is motion) for the development in this chapter.

Let's examine the geometry first. Instead of a trial-and-error of the last example, we use the following algorithm. We build a *sequence of secant lines*. Each line passes through two points, the point $A = (a, f(a))$ (same for all) and some variable point $B = (b, f(b))$ with $a \neq b$:

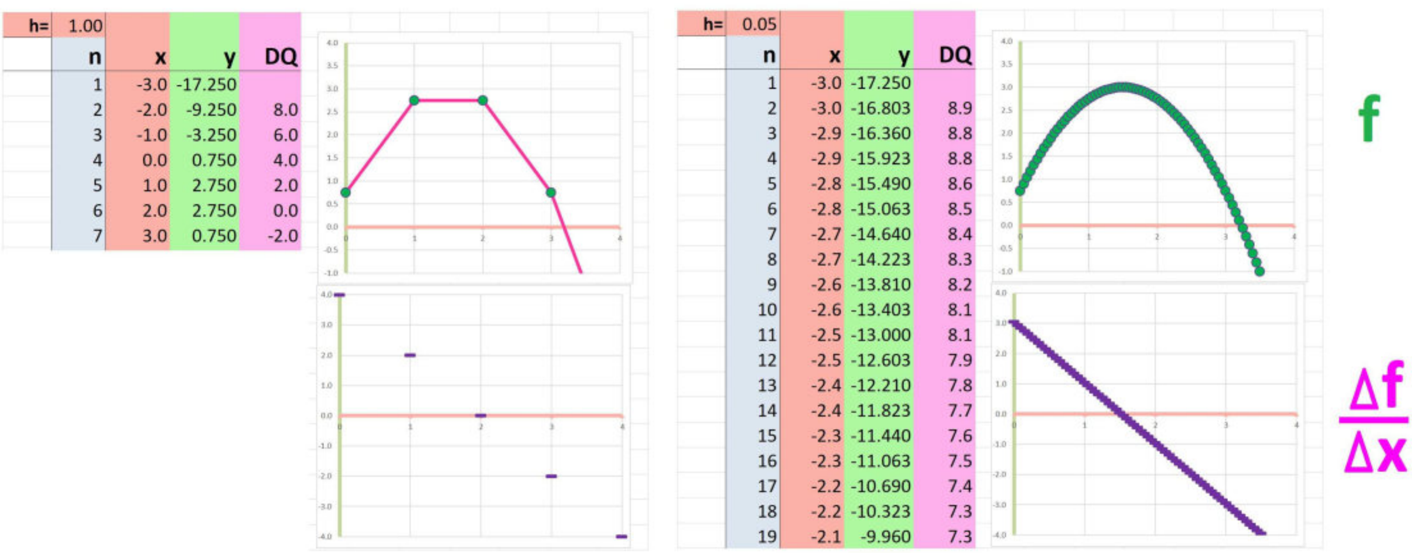


These secants are expected to get closer and closer to the tangent as B is getting closer and closer to A.

However, the secants are also becoming shorter and shorter, and less and less useful. If we extend these segments into straight lines, we can observe them *rotate*. They will rotate less and less as they are getting closer and closer to the tangent line.

But an even better idea is algebraic: Follow the *slopes*! Since they are nothing but numbers, we are dealing with a *sequence*. A sequence and its *limit*!

We will, however, set limits aside for now and address first the other challenge posed above: how to deal with different slopes at different locations, i.e., the *difference quotient*:



h=	1.00		
n	x	y	DQ
1	-3.0	-17.250	
2	-2.0	-9.250	8.0
3	-1.0	-3.250	6.0
4	0.0	0.750	4.0
5	1.0	2.750	2.0
6	2.0	2.750	0.0
7	3.0	0.750	-2.0

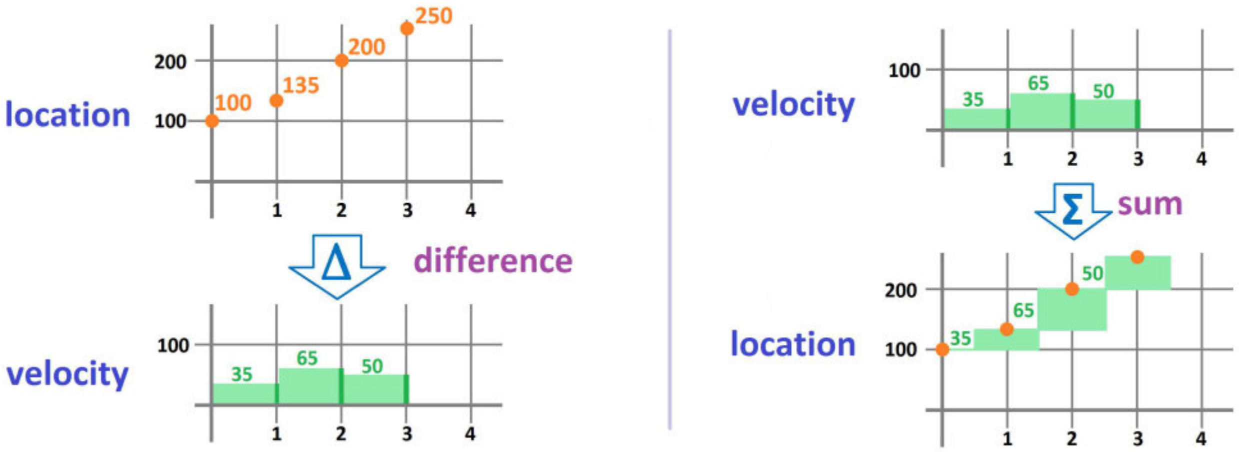
h=	0.05		
n	x	y	DQ
1	-3.0	-17.250	
2	-3.0	-16.803	8.9
3	-2.9	-16.360	8.8
4	-2.9	-15.923	8.8
5	-2.8	-15.490	8.6
6	-2.8	-15.063	8.5
7	-2.7	-14.640	8.4
8	-2.7	-14.223	8.3
9	-2.6	-13.810	8.2
10	-2.6	-13.403	8.1
11	-2.5	-13.000	8.1
12	-2.5	-12.603	7.9
13	-2.4	-12.210	7.8
14	-2.4	-11.823	7.7
15	-2.3	-11.440	7.6
16	-2.3	-11.063	7.5
17	-2.2	-10.690	7.4
18	-2.2	-10.323	7.3
19	-2.1	-9.960	7.3

Any computation of the slope starts with a computation of the *rise*.

3.2. The difference of a sequence and the difference of a function

We enter calculus through a study of motion.

Recall from Chapter 1 our running example of a broken speedometer and a broken odometer. When the intervals of time are units, we can go from locations to velocities and back with the two simple operations of *difference and sum*:



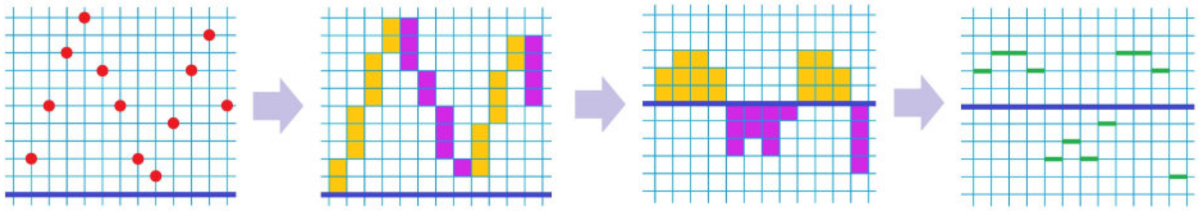
This is the summary:

- If each term of a sequence represents a location, the pair-wise differences will give you the *velocities*. A broken speedometer is substituted with an odometer and a watch.
- If each term of a sequence represents a velocity, their sum up to that point will give you the *location*. A broken odometer is substituted with a speedometer and a watch.

In the abstract, the pairwise differences represent the *change* within the sequence, from each of its terms to the next (seen in Volume 1, Chapter 1PC-1).

Example 3.2.1: sequence given by graph

In the simplest case, a sequence takes only integer values, then on the graph of the sequence, we just count the number of steps we make, up and down:



These increments then make a new sequence plotted on the right.

The idea leads to the following definition:

Definition 3.2.2: sequence of differences

For a sequence a_n, its *sequence of differences*, or simply the difference, is a new sequence, say d_n, defined for each n by the following:

$$d_n = a_{n+1} - a_n \,.$$

It is denoted as follows:

$$\boxed{\Delta a_n = a_{n+1} - a_n}$$

Those are the *rises* that we will soon be using to compute some slopes.

This is how the new sequence is built:

Sequence of differences

$$\begin{array}{rccccccccccc}
\text{a sequence:} & a_1 & & & & a_2 & & & & a_3 & & & & a_4 & ...\\
 & & \searrow & & \swarrow & & \searrow & & \swarrow & & \searrow & & \swarrow & & ...\\
\text{its differences:} & & & a_2-a_1 & & & & a_3-a_2 & & & & a_4-a_3 & & & ...\\
 & & & || & & & & || & & & & || & & & ...\\
\text{a new sequence:} & & & d_1 & & & & d_2 & & & & d_3 & & & ...\\
 & & & || & & & & || & & & & || & & & ...\\
\text{the notation:} & & & \Delta a_1 & & & & \Delta a_2 & & & & \Delta a_3 & & & ...
\end{array}$$

Example 3.2.3: sequence given by list

When a sequence is given by a list, we subtract the last term from the current one and put the result below as follows:

$$\begin{array}{rcccccccccccccc}
\text{a sequence:} & 2 & & & & 4 & & & & 7 & & & & 1 & & ...\\
 & & \searrow & & \swarrow & & \searrow & & \swarrow & & \searrow & & \swarrow & & \searrow & ...\\
\text{its differences:} & & & 4-2 & & & & 7-4 & & & & 1-7 & & & & ...\\
 & & & || & & & & || & & & & || & & & & ...\\
\text{a new sequence:} & & & 2 & & & & 3 & & & & -6 & & & & ...
\end{array}$$

We have a new list.

Exercise 3.2.4

Plot the location and the velocity for the following trip: "I drove fast, then gradually slowed down, stopped for a very short moment, gradually accelerated, maintained speed, hit a wall." Make up your own story and repeat the task.

Exercise 3.2.5

Draw a curve on a piece of paper, imagine that it represents your locations, and then sketch what your velocity would look like. Repeat.

Exercise 3.2.6

Imagine that the first graph represents, instead of locations, the balances of bank accounts. Describe what has been happening.

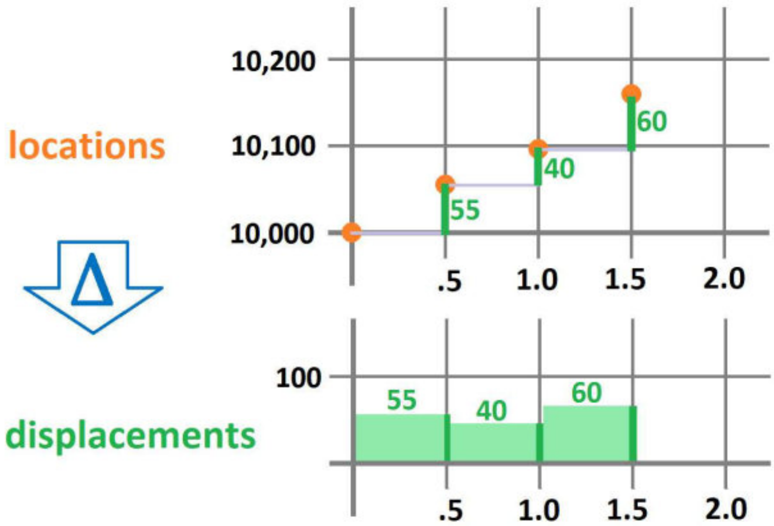
Example 3.2.7: velocities

Now, what if the time intervals aren't units?

We still look at the odometer *several* times during the trip and record the mileage on a piece of paper. But this time it's every half-hour. This list of our consecutive *locations* might look like this:

- initial reading: $10,000$ miles
- after the first half-hour: $10,055$ miles
- after the second half-hour: $10,095$ miles
- after the third half-hour: $10,155$ miles
- etc.

We plot this data (top):



We also compute the differences (bottom).

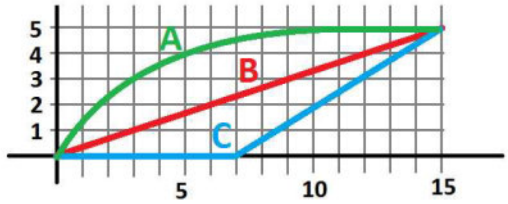
What is the meaning of the differences? They are the distances covered during each of these half-hour periods, by *subtraction*:

- distance covered during the first half-hour: $10,055 - 10,000 = 55$ miles
- distance covered during the second half-hour: $10,095 - 10,055 = 40$ miles
- distance covered during the third half-hour: $10,155 - 10,095 = 60$ miles
- etc.

They are called the *displacements*.

Example 3.2.8: three runners

The graph below shows the positions of three runners in terms of time, n. Describe what has happened:



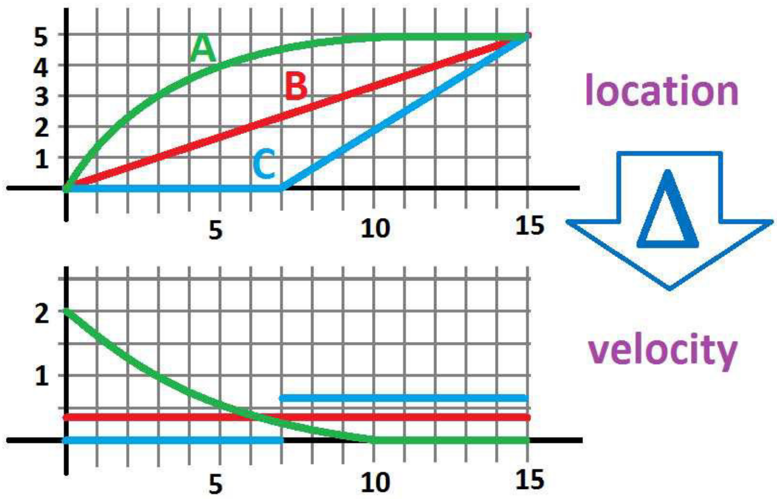
They are all at the starting line together, and at the end, they are all at the finish line. Furthermore, A reaches the finish line first, followed by B, and then C (who also starts late).

This is *how* each did it:

- A starts fast, then slows down, and almost stops close to the finish line.
- B maintains the same speed.
- C starts late and then runs fast at the same speed.

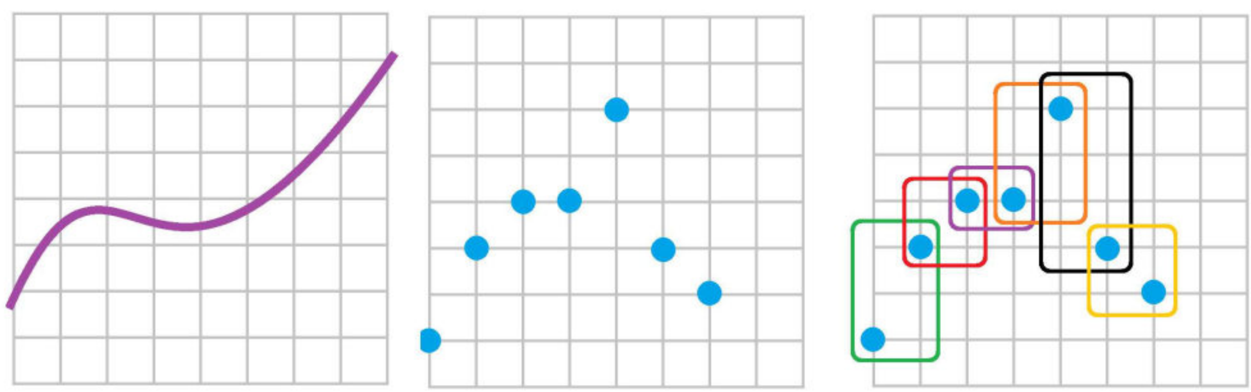
We can see that A is running faster because the distance from B is increasing. It becomes slower later, which is visible from the decreasing distance.

There is another way! We can discover this and the rest of the facts by examining the graphs of the *differences* of the sequences:



For example, we clearly see the constant velocities of the three at the end of the run.

In general, we face *functions*:

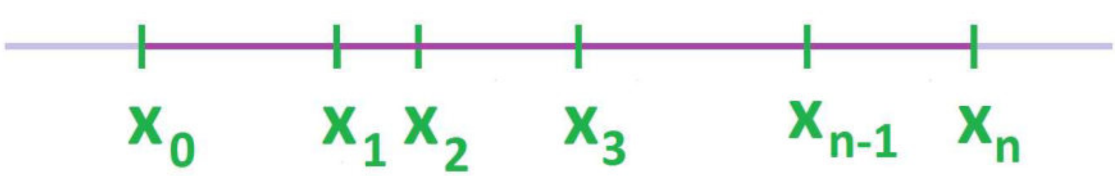


In order to compute the difference of a function, we need to *sample* it. This is how it is done.

Suppose we have an interval $[a, b]$. We construct its *partition* as follows. First, we choose a natural number n and then place $n + 1$ points on the interval:

$$a = x_0 < x_1 < x_2 < ... < x_{n-1} < x_n = b\,.$$

In other words, we have a strictly increasing sequence x_i:



As a result, the interval is split into n smaller intervals of possibly different lengths:

$$[x_0, x_1],\ [x_1, x_2],\ ...,\ [x_{n-1}, x_n]\,.$$

We will use the following terminology:

Definition 3.2.9: partition of an interval

A *partition of an interval* $[a, b]$ is its representation as the union of intervals that intersect only at their end-points:

$$[a, b] = [x_0, x_1] \cup [x_1, x_2] \cup ... \cup [x_{n-1}, x_n]\,.$$

These end-points,

$$a = x_0,\ x_1,\ x_2,\ ...,\ x_{n-1},\ x_n = b\,,$$

will be called the *nodes* of the partition.

The difference of this sequence gives us the lengths of these intervals:

Definition 3.2.10: increments of partition

The *increments* of the partition are given by:

$$\Delta x_i = x_i - x_{i-1},\ i = 1, 2, ..., n\,.$$

It is simply Δx when they are equal.

Those are the *runs* that we will soon be using to compute some slopes.

Example 3.2.11: description of a drive

Suppose we drove around while paying attention to the clock and to the mileposts. We produced this simple table with *five* columns:

time (hours):	0	2	4	6	8
location (miles):	0	60	160	80	0

What were the displacements over these *four* periods of time? We use the difference formula:

$$\boxed{\text{displacement} \quad = \quad \text{change of location}}$$

These are the computations:

time (hours):	0	2	4	6	8
location (miles):	0	60	160	80	0
displacement (miles):	$60 - 0 =$	60			
displacement (miles):		$160 - 60 =$	100		
displacement (miles):			$80 - 160 =$	-80	
displacement (miles):				$0 - 80 =$	-80

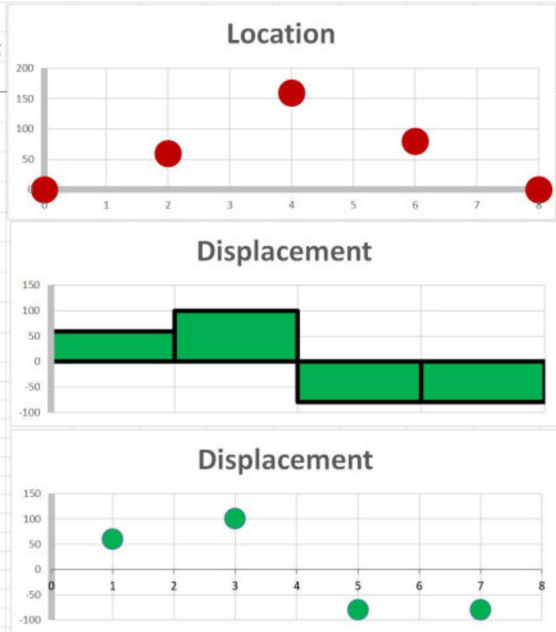
Because the time intervals are $\Delta x = 1$, these displacements are also the *velocities*.

The four computed values are the displacements over the following *intervals* of time: $[0, 2]$, $[2, 4]$, $[4, 6]$, and $[6, 8]$, respectively. The result is this table:

time intervals (hours):	$[0, 2]$	$[2, 4]$	$[4, 6]$	$[6, 8]$
velocity (miles/hour):	60	100	-80	-80

This is the summary of what we have done:

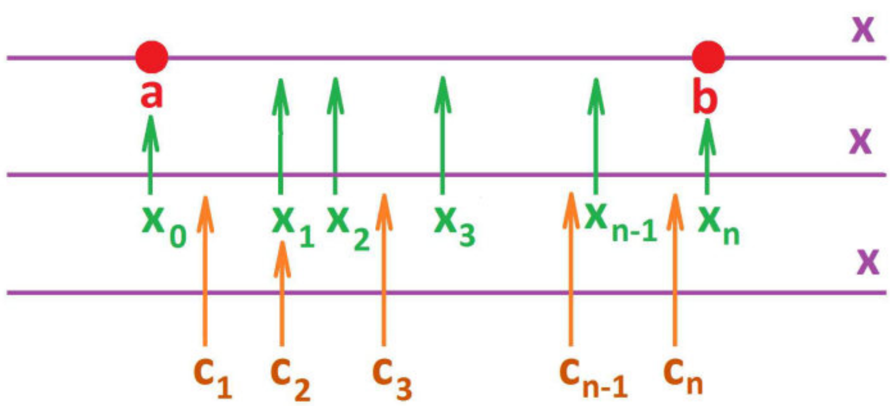
time	location	time	displacement
h	m	h	m
0	0		
1		1	60
2	60		
3		3	100
4	160		
5		5	-80
6	80		
7		7	-80
8	0		



We can also choose to assign the four values to, say, the *middle points* of these intervals, as follows:

time (hours):	1	3	5	7
velocity (miles/hour):	60	100	-80	-80

So, in addition to the nodes, the *primary nodes*, we may also be given the *secondary nodes* in each interval of the partition:



We have another sequence:

$$c_1 \text{ in } [x_0, x_1],\ c_2 \text{ in } [x_1, x_2],\ ...,\ c_n \text{ in } [x_{n-1}, x_n]\,.$$

The definition below summarizes this setup:

Definition 3.2.12: augmented partition of interval

An *augmented partition*, or simply a partition, of an interval $[a, b]$ consists of two sequences:

1. *primary nodes* $a = x_0,\ x_1,\ x_2,\ ...,\ x_{n-1},\ x_n = b$
2. *secondary nodes* $c_1,\ c_2,\ c_3,\ ...,\ c_{n-1},\ c_n$

that satisfy these inequalities:

$$x_0 \le c_1 \le x_1 \le c_2 \le x_2 \le ... \le x_{n-1} \le c_n \le x_n\,.$$

Warning!

We can choose secondary nodes from the list of primary nodes because the inequalities are non-strict.

A partition is a proper setting for sampling functions.

Example 3.2.13: sampling

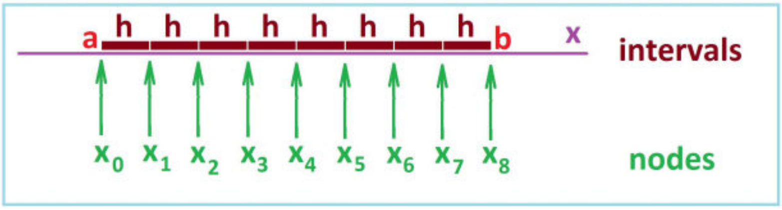
Let's generalize the last example. First, we consider the trip with a broken speedometer in a more general setting. We have a time interval $[a,b]$. In order to estimate our speed, we decide to look at the milestones several times during the trip. The moments of time may be random: $a = x_0,\ x_1,\ x_2,\ ...,\ x_{n-1},\ x_n = b$. In other words, we have a partition of $[a,b]$, and we sample the function f that represents our location at the primary nodes of the partition. We have our answer but we can also assign these numbers to the secondary nodes as a matter of bookkeeping:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
time (hours):	c_1	c_2	c_3	c_4
velocity (miles/hour):	60	100	-80	-80

In the examples, we chose the secondary nodes in a consistent way. First we can choose equal increments:

$$\boxed{h = \Delta x = \frac{b-a}{n}}$$

With this assumption, we now choose the secondary nodes:



There are three main "schemes" for choosing secondary nodes. One is seen above: The secondary nodes are placed at the end of each interval. It is called the *right-end scheme*:

- The primary nodes are $x = a, a+h, a+2h, ...$.
- The secondary nodes are $c = a+h, a+2h, ...$.

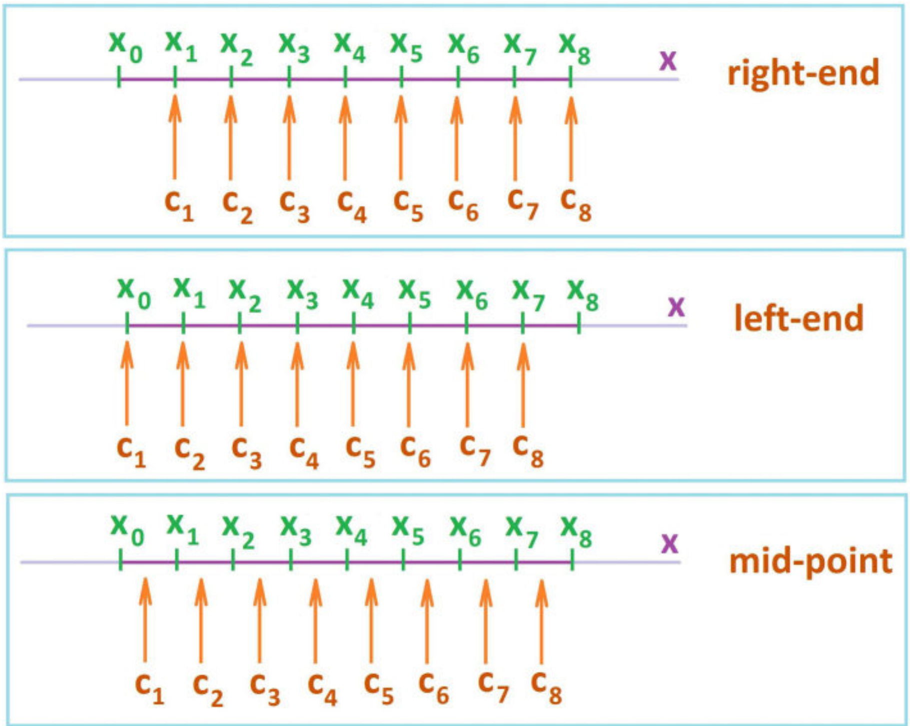
This is the *left-end scheme*:

- The primary nodes are $x = a, a+h, a+2h, ...$.
- The secondary nodes are $c = a, a+h, ...$.

Another convenient choice is the *mid-point scheme*:

- The primary nodes are $x = a, a+h, a+2h, ...$.
- The secondary nodes are $c = a+h/2, a+3h/2, ...$.

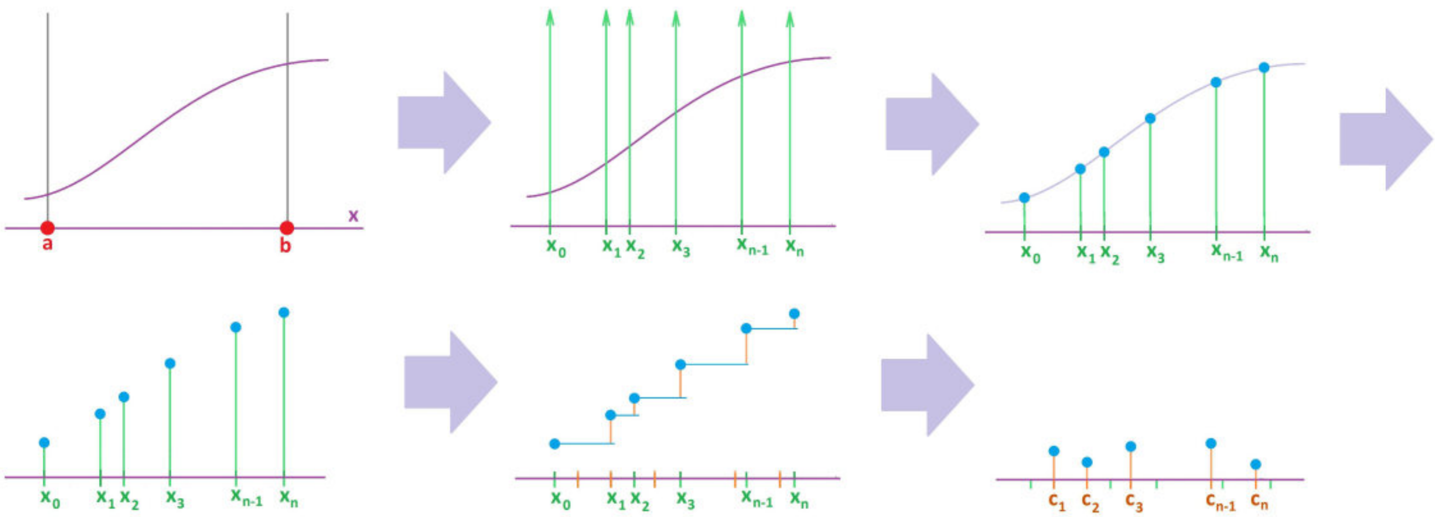
They are illustrated below:



The final step is the following:

► We utilize the secondary nodes as the inputs of a new function.

The whole construction is outlined below:



These are the stages that we see here:

1. a function,
2. a partition and its primary nodes,
3. sampling of the function at the primary nodes,
4. removing the graph,
5. plotting the differences,
6. placing these differences at the secondary nodes.

The result is a new function.

It is defined algebraically as follows:

Definition 3.2.14: difference of function

Suppose $y = f(x)$ is defined at the nodes x_k, $k = 0, 1, 2, ..., n$, of a partition. Then the *difference* of f is a function defined at the secondary nodes of the partition, and denoted, as follows:

$$\boxed{\Delta f(c_k) = f(x_{k+1}) - f(x_k)}$$

Example 3.2.15: squaring function

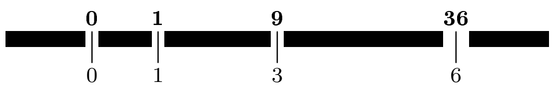
We let $f(x) = x^2$ and choose a partition of $[0, 6]$:

- The primary nodes are $0, 1, 3, 6$.
- The secondary nodes are $0, 2, 4$.

We sample the function:

$$f(0) = 0, \ f(1) = 1, \ f(3) = 9, \ f(6) = 36\,.$$

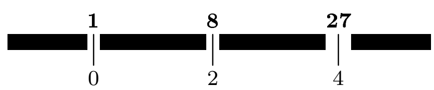
This is the sampled function:



Then

$$\begin{array}{llll} \Delta f(0) & = f(1) - f(0) & = 1^2 - 0^2 & = 1 \\ \Delta f(2) & = f(3) - f(1) & = 3^2 - 1^2 & = 8 \\ \Delta f(4) & = f(6) - f(3) & = 6^2 - 3^2 & = 27 \end{array}$$

This is the difference of the sampled function:



If we change the secondary nodes, we, of course, change the differences. However, the results are the same on a large scale, as we shall see later.

The following language will be often used to illustrate these mathematical ideas:

Definition 3.2.16: displacement

When a function or a sequence is called "location" or "position", its difference is called the *displacement*.

When a partition is specified, we may omit the subscript for the nodes, x, and the secondary nodes, c. Then we can use the following simplified notation:

Difference

$$\Delta f(c) = f(x + \Delta x) - f(x)$$

How do we treat motion when the time increment isn't 1? How do we find the *velocities* over time? The difference construction presented in this section paves the way: Whenever we subtract, we also divide.

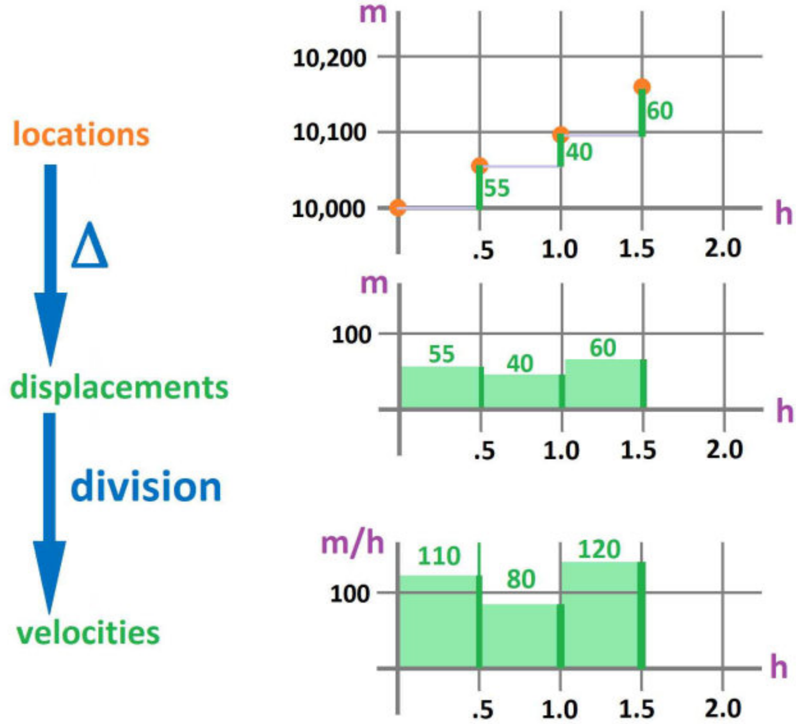
Exercise 3.2.17

You have received the following email from your boss: "Tim, Look at the numbers in this spreadsheet. This stock seems to be inching up... Does it? If does, how fast? Thanks. – Tom". Describe your actions.

3.3. The rate of change: the difference quotient

We have been computing the *rises*. To get the slope from those, we just divide by the *runs*.

Once again let's consider the running example of a *broken speedometer* from Chapter 1. Just as in the last section, we find the displacement for each interval of time, which is half-hour long. The new step is the division of the differences by the time increment $h = .5$ producing the velocities:



Example 3.3.1: velocity

We now consider a more elaborate version of the problem of *incremental motion*.

Suppose we drove around while paying attention to the clock and to the mileposts. The result is this simple table with *five* columns:

time (hours):	0	2	4	6	8
location (miles):	0	60	160	80	0

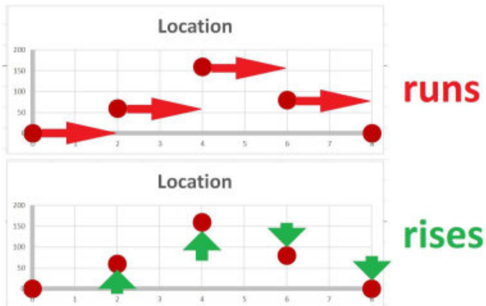
What was the velocity over these *four* periods of time? We estimate it by the "difference quotient" formula:

$$\boxed{\text{velocity} = \frac{\text{change of location}}{\text{change of time}}}$$

These are the computations:

time (hours):	0	2	4	6	8
location (miles):	0	60	160	80	0
velocity (miles/hour):	$\frac{60-0}{2-0} =$	30			
velocity (miles/hour):		$\frac{160-60}{4-2} =$	50		
velocity (miles/hour):			$\frac{80-160}{6-4} =$	-40	
velocity (miles/hour):				$\frac{0-80}{8-6} =$	-40

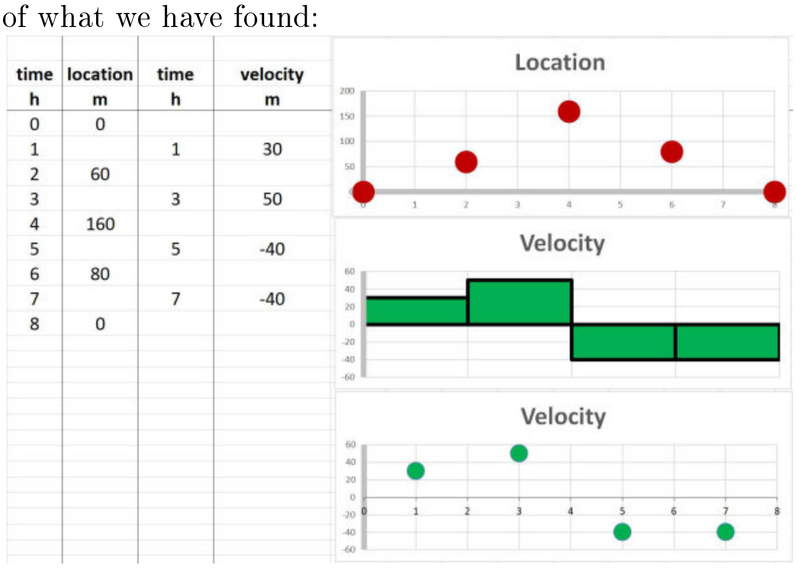
The numerators and denominators are nothing but the rises and runs:



The four computed values are the average velocities over the following *intervals* of time: $[0,2]$, $[2,4]$, $[4,6]$, and $[6,8]$, respectively. The result is this table:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
velocity (miles/hour):	30	50	-40	-40

This is the summary of what we have found:



time	location	time	velocity
h	m	h	m
0	0		
1		1	30
2	60		
3		3	50
4	160		
5		5	-40
6	80		
7		7	-40
8	0		

We may also choose to assign the four values to the secondary nodes (the middle points are shown in the last chart), as follows:

time (hours):	1	3	5	7
velocity (miles/hour):	30	50	-40	-40

In general, we see the time and the location as just two separate sequences, say, x_n and y_n. Then the velocity is the increment of the latter over the increment of the former. We notice that those two are the differences of the two sequences:

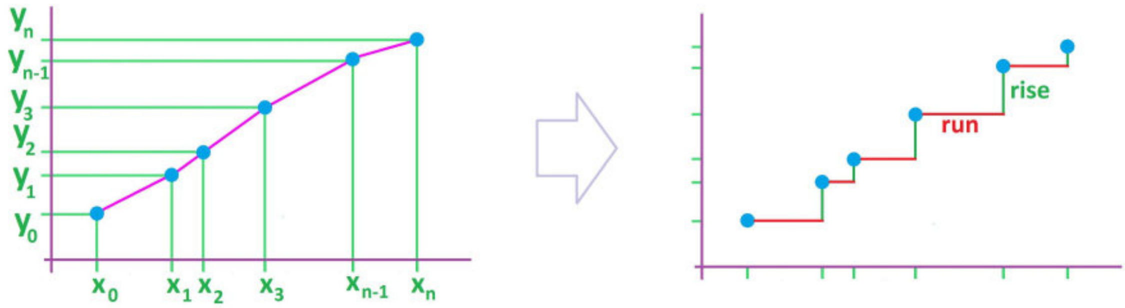
Definition 3.3.2: difference quotient for sequences

The *difference quotient* of a sequence y_n with respect to a sequence x_n is defined to be the sequence that is the difference of y_n divided by the difference of x_n:

$$\boxed{\frac{\Delta y_n}{\Delta x_n} = \frac{y_{n+1} - y_n}{x_{n+1} - x_n}}$$

provided the denominator is not zero.

It is the relative change – the *rate* of change – of the two sequences. And for each consecutive pair of points, it is the slope of the line connecting them:



This is why it is called this way:

Difference quotient

difference, subtraction

$$\frac{\Delta y_n}{\Delta x_n} = \frac{y_{n+1} - y_n}{x_{n+1} - x_n} \quad \swarrow \quad \longleftarrow \quad \nwarrow$$

quotient, division

difference, subtraction

Example 3.3.3: velocity, continued

Let's continue with the last example. This time, we also proceed to compute the *acceleration*. Just as we used the "difference quotient" formula to find the velocity from the location, we now use it to find the acceleration from the velocity:

$$\boxed{\text{acceleration} = \frac{\text{change of velocity}}{\text{change of time}}}$$

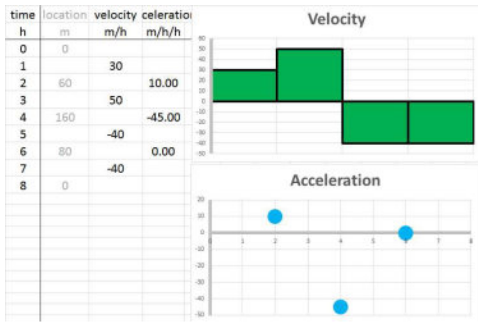
We apply this formula to *three* periods of time. These are the computations:

time intervals (hours):	$[0,2]$	$[2,4]$	$[4,6]$	$[6,8]$
time (hours):	1	3	5	7
velocity (miles/hour):	30	50	-40	-40
acceleration (miles/hour/hour):	$\frac{50-30}{3-1} =$	10		
acceleration (miles/hour/hour):		$\frac{-40-50}{5-3} =$	-45	
acceleration (miles/hour/hour):			$\frac{-40-(-40)}{7-5} =$	0

The three computed values are the average accelerations over the following *intervals* of time: $[0,4]$, $[2,6]$, and $[4,8]$, respectively. This is the result:

time intervals (hours):	$[0,4]$	$[2,6]$	$[4,8]$
acceleration (miles/hour/hour):	10	-45	0

This is the summary of what we have found:



Alternatively, we may choose to assign the three values to the *middle points* of these intervals, as follows:

time (hours):	2	4	6
acceleration (miles/hour/hour):	10	-45	0

These happen to be the primary nodes of our partition!

Example 3.3.4: computing motion

In general, when the location is known for numerous moments of time, we have these sequences:

- t_n for the time,
- p_n for the position,
- v_n for the velocity, and
- a_n for the acceleration.

They are connected to each other by these formulas:

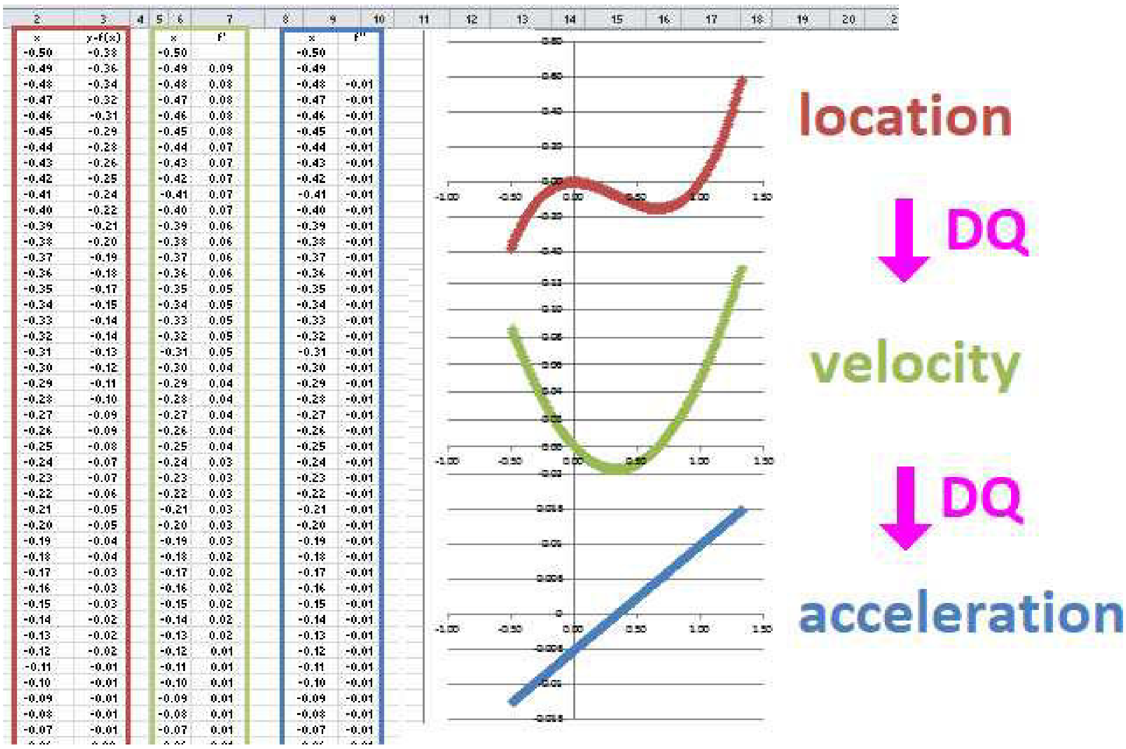
$$v_{n+1} = \frac{p_{n+1} - p_n}{t_{n+1} - t_n} \quad \text{and} \quad a_{n+1} = \frac{v_{n+1} - v_n}{t_{n+1} - t_n}\,.$$

Both are nothing but the difference quotient!

For computations with a lot of data, a *spreadsheet* is used. The two formulas have the same form:

```
=(RC[-1]-R[-1]C[-1])/(RC2-R[-1]C2)
```

The formula for the velocity (or the acceleration respectively) refers to the column that contains the location (or velocity respectively) in the numerator and to the column that contains the time in the denominator. Possible data and graphs for these three are shown below:



The graphs are made of dots only! The second column (velocity) has one fewer data point and the next (acceleration) one fewer yet. However, when zoomed out, the graphs give the impression that the three functions have the same domain.

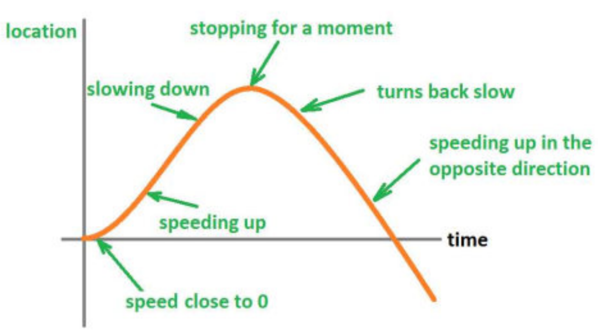
The story of what happened can be inferred from these graphs:

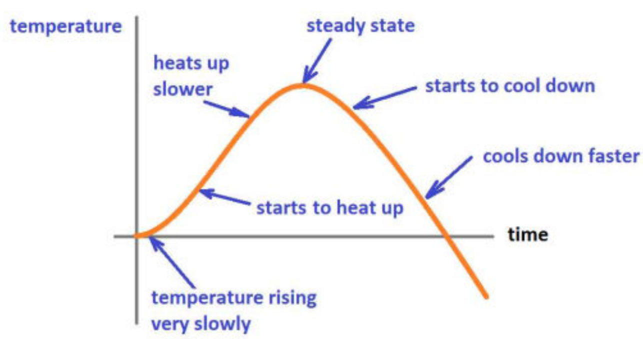
1. The object was moving forward, then stopped for a moment, turned back for a short while, then started moving forward again.
2. The object velocity was forward high, then lower and lower (slower and slower), until it was zero and then changed direction, then forward again, higher and higher (faster and faster).
3. The object had negative acceleration (deceleration), which then became positive, and continued to grow.

However, it is much easier to draw conclusions about the velocity from its graph than from the graph of the position!

These formulas will allow us to develop realistic models of motion.

We can continue to increase the number of data points and, as we zoom out, the scatter plots will look like *continuous curves*! The analysis presented in this section remains fully applicable. It amounts to looking at how fast the vertical location is changing relative to the change of the horizontal location (left):





Furthermore, we replace (right) our time-dependent quantity, location, for another, temperature as just one of the examples of the breadth of applicability of these ideas. This dictates the need for context-independent, mathematical terminology. We use the following:

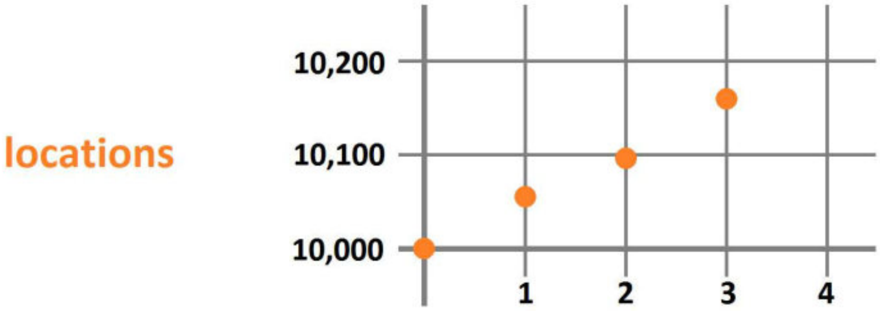
- The difference quotient is the *rate of change* of the output per change of the input.

Next, the sequences may come from *sampled functions*. A sequence of inputs will produce a sequence of

outputs:

$$a_n = f(x_n)\,.$$

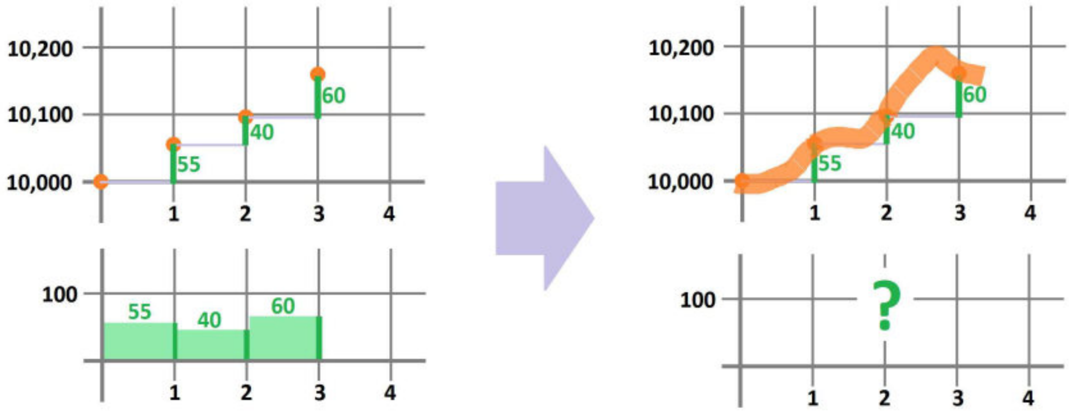
We approached the problem by plotting the location as a function of time:



We then found:

$$\text{average velocity} \;=\; \frac{\text{displacement}}{\text{time}}$$

for each of the time periods (left):



However, what if behind this data is a continuously changing location (right)? With so much information, can we find the "exact" velocity for all of these moments of time?

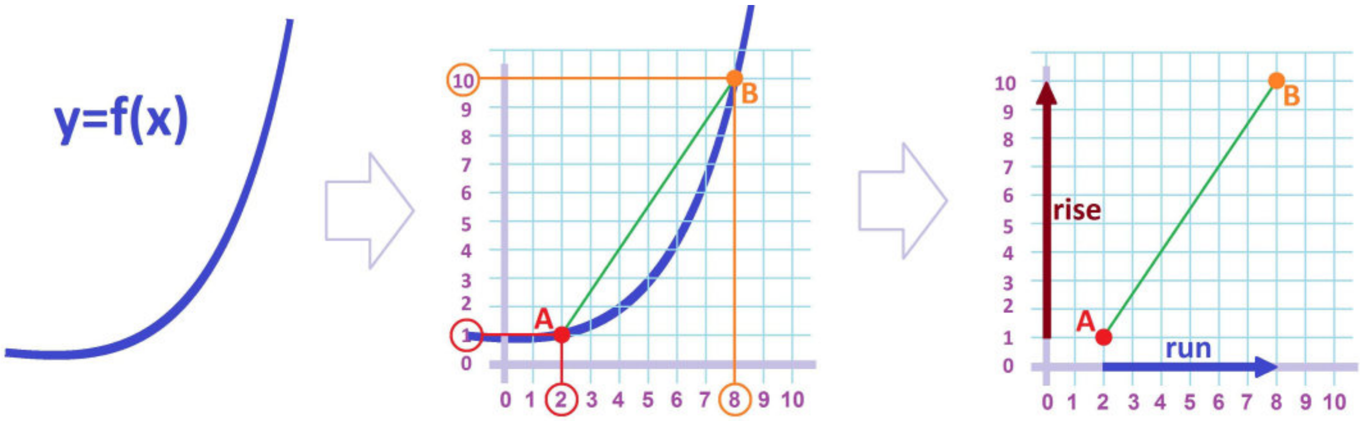
Suppose we know only *two* values of a function:

$$f(x_1) = y_1 \quad \text{and} \quad f(x_2) = y_2\,,$$

with $x_1 \neq x_2$. Then, what can we say about its rate of change? The answer is given by the *slope* of the line through these two points on the graph of $y = f(x)$:

$$A = (x_1, y_1) \quad \text{and} \quad B = (x_2, y_2)\,,$$

as follows:



The slope is, of course, the rise over the run:

$$\text{slope } = \frac{y_2 - y_1}{x_2 - x_1} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}\,.$$

The numerator, the rise, is the change of y, which is also the difference of f,:

$$\Delta f = f(x_2) - f(x_1)\,.$$

The denominator, the run, is the change of x, which we will call the *increment* of x:

$$\Delta x = x_2 - x_1\,.$$

Their ratio, the slope of the line, is the rate of change of f, the difference quotient of f:

$$\frac{\Delta f}{\Delta x} = \frac{f(x_2) - f(x_1)}{x_2 - x_1}$$

Example 3.3.5: single segment

In the above picture, we have:

1. The increment of x is $\Delta x = x_2 - x_1 = 8 - 2 = 6\,.$
2. The difference of f is $\Delta f = f(x_2) - f(x_1) = 10 - 1 = 9\,.$
3. The difference quotient of f is $\dfrac{\Delta f}{\Delta x} = \dfrac{9}{6}\,.$

Example 3.3.6: speed

Suppose we travel by car and record our location several times (first row):

$$\boxed{\begin{array}{ccccc} - & 100 & - - - & 250 & - \\ - & -\bullet- & - - - & -\bullet- & - \\ & 0 & & 2 & \end{array}} \longrightarrow \boxed{\begin{array}{ccccc} - & 100 & - - - & 250 & - \\ - & -\bullet- & -\ 75\ - & -\bullet- & - \\ & 0 & 1 & 2 & \end{array}}$$

We then compute the velocity (i.e., difference quotient) during each segment of time and place the number in between (second row).

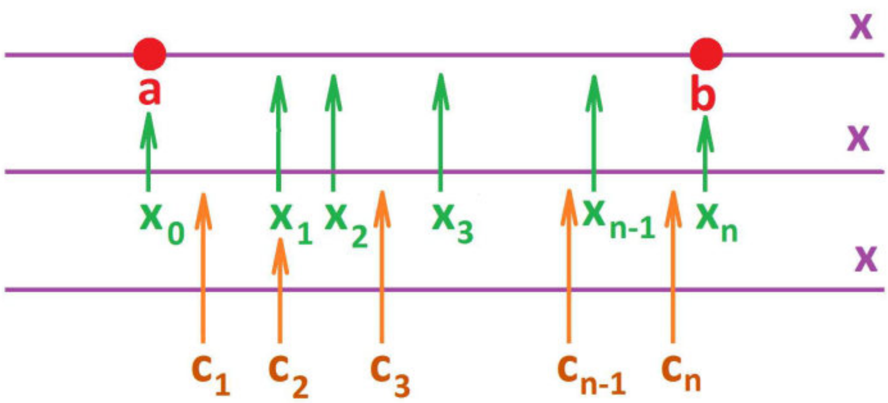
By analogy, if we know only *two* values of a function (first row) at ends of an interval, we compute the difference quotient along this interval (second row):

$$\begin{array}{ccccc} - & f(x) & - - - & f(x+\Delta x) & - \\ - & -\bullet- & \dfrac{\Delta f}{\Delta x} & -\bullet- & - \\ & x & c & x+\Delta x & \end{array}$$

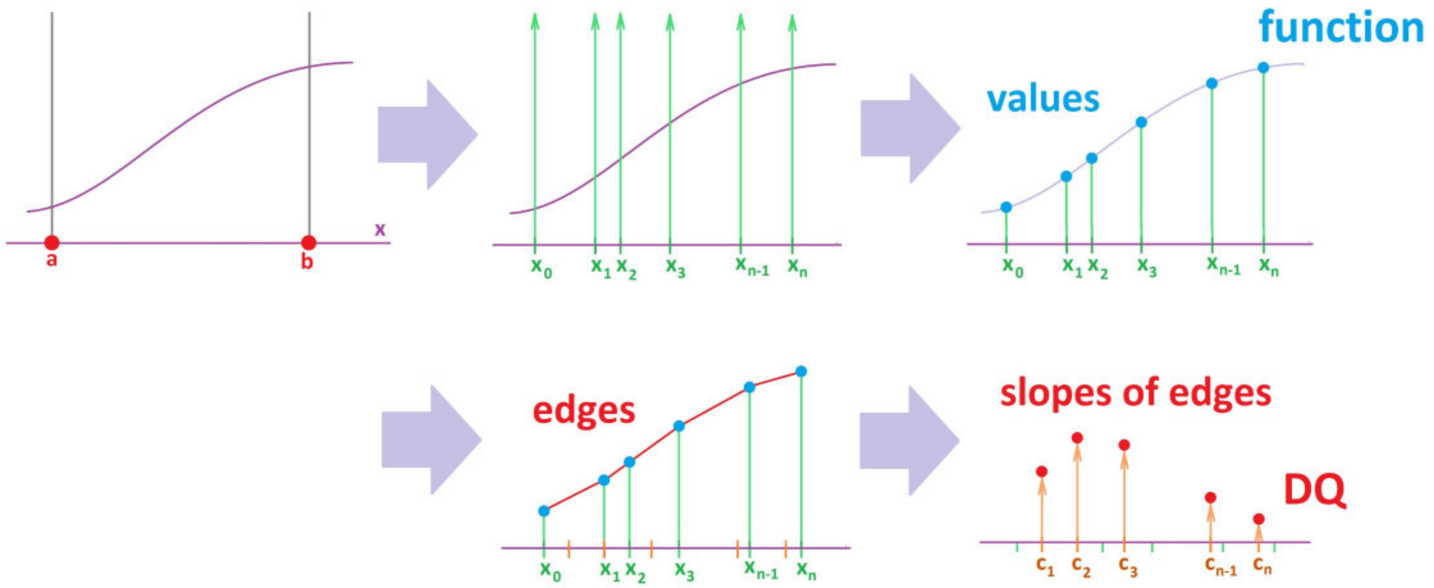
Now, what if a function $y = f(x)$ is known for *several* values of x within an interval $[a, b]$? We follow the idea of the construction in the beginning of the chapter.

Suppose we have constructed an *augmented partition* an interval $[a, b]$ as in the last section:

$$a = x_0 \le c_1 \le x_1 \le c_2 \le x_2 \le \ldots \le x_{n-1} \le c_n \le x_n = b\,.$$



Just as in the last section, we utilize the secondary nodes as the inputs of the new function:



However, this time we are not using the differences of values (the heights of the steps) but the *slopes* of the inclines. We define the most fundamental operation of calculus below:

Definition 3.3.7: difference quotient for functions

Suppose a function $y = f(x)$ is defined at the primary nodes $x_k,\ k = 0, 1, 2, ..., n$, of a partition. Then the *difference quotient* of f is defined at the secondary nodes $c_k,\ k = 1, 2, ..., n$ of the partition as this fraction:

$$\boxed{\frac{\Delta f}{\Delta x}(c_k) = \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k} = \frac{f(x_k + \Delta x_k) - f(x_k)}{\Delta x_k}}$$

When a partition is specified, we may omit the subscript for the nodes, x, and the secondary nodes, c. Then we can use the following simplified notation:

$$\frac{\Delta f}{\Delta x}(c) = \frac{f(x + \Delta x) - f(x)}{\Delta x}$$

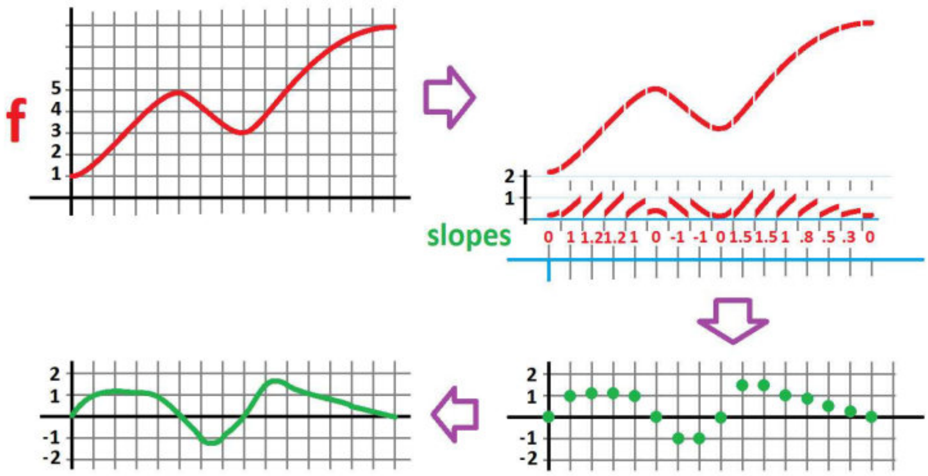
Below is the breakdown of the notation:

Difference quotient

old function
$\downarrow$
a fraction: $\dfrac{\Delta f}{\Delta x}\ (c)$
$\nearrow$ $\uparrow$
new function input variable

Example 3.3.8: graphs

The difference is a new sequence. Can we plot it based on the *graph* of f only? Yes, but not at once. We do this piece by piece. We take the graph of the function (top left) and, at many locations, cut – using the grid – a segment of the graph so short that it's *almost* straight (top right):



The vertical locations of these pieces are irrelevant. We line them up below so that their slopes, the difference quotients of the function, are easy to estimate (bottom right). This gives us 15 numbers. These points can be connected to form a curve (bottom left). This looks like a new function!

Example 3.3.9: tables

Now we plot $\frac{\Delta f}{\Delta x}$ based on the *values* of f with the increment of $\Delta x = .5$:

x	0	$.5$	1.0	1.5	2.0
$y = f(x)$	-1	-2	0	1	1

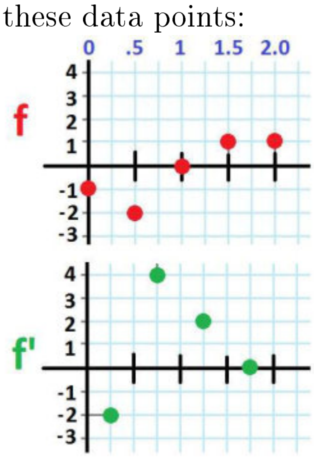
We compute the difference quotient from this data:

$$\frac{\Delta f}{\Delta x}(x) = \frac{f(x + \Delta x) - f(x)}{\Delta x} .$$

We do that interval by interval:

$[x, x + \Delta x]$	$[0, .5]$	$[.5, 1.0]$	$[1.0, 1.5]$	$[1.5, 2.0]$
$\Delta f\,(c) = f(x + \Delta x) - f(x) =$	$-2 - (-1)$	$0 - (-2)$	$1 - 0$	$1 - 1$
$=$	-1	2	1	0
$\frac{\Delta f}{\Delta x}(c) =$	$-1/.5$	$2/.5$	$1/.5$	$0/.5$
$=$	-2	4	2	0

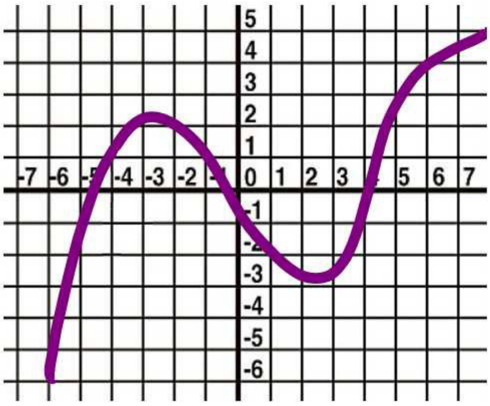
The results are confirmed by plotting these data points:



These numbers are, of course, just the slopes of the lines that connect the dots.

Exercise 3.3.10

The graph of a function f is given below:



Suppose we have a partition with:

$$[a,b]=[0,4] \quad\text{and}\quad \Delta x=1\,.$$

Estimate the values of the difference quotient $\dfrac{\Delta f}{\Delta x}$. Show your constructions and computations.

Example 3.3.11: spreadsheet

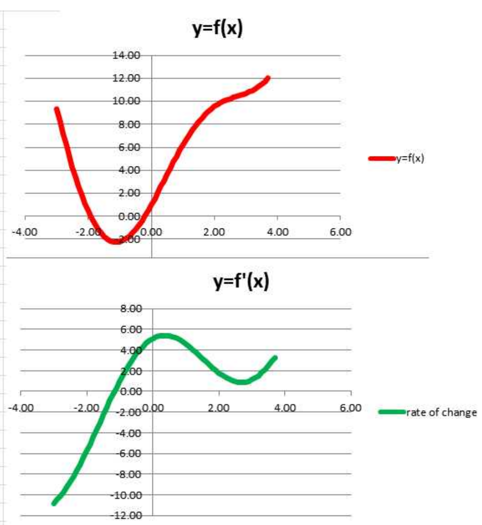
We now utilize a spreadsheet to speed up this process. The function is given by a possibly large table of values, with two columns: x and $y=f(x)$. Then, for each pair of consecutive values, we compute:

- the increment of x,
- the increment of y, and
- their ratio, the difference quotient.

The last formula is:

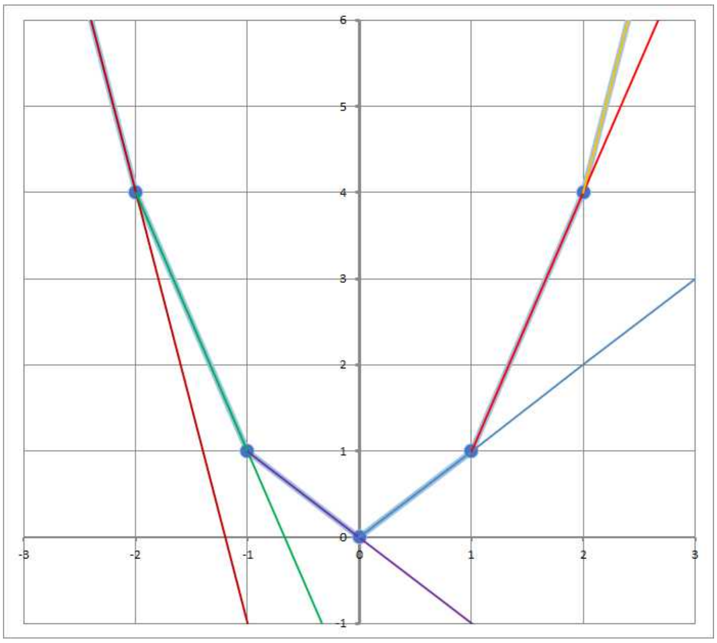
```
=RC[-2]/RC[-3]
```

x	y=f(x)	change of x	change of y	x	rate of change
-3.00	9.29	0.10	-1.08	-3.00	-10.81
-2.90	8.21	0.10	-1.05	-2.90	-10.49
-2.80	7.17	0.10	-1.01	-2.80	-10.12
-2.70	6.15	0.10	-0.97	-2.70	-9.71
-2.60	5.18	0.10	-0.92	-2.60	-9.25
-2.50	4.26	0.10	-0.87	-2.50	-8.75
-2.40	3.38	0.10	-0.82	-2.40	-8.21
-2.30	2.56	0.10	-0.76	-2.30	-7.64
-2.20	1.80	0.10	-0.70	-2.20	-7.04
-2.10	1.09	0.10	-0.64	-2.10	-6.40
-2.00	0.45	0.10	-0.58	-2.00	-5.75
-1.90	-0.12	0.10	-0.51	-1.90	-5.08
-1.80	-0.63	0.10	-0.44	-1.80	-4.39
-1.70	-1.07	0.10	-0.37	-1.70	-3.70
-1.60	-1.44	0.10	-0.30	-1.60	-3.00
-1.50	-1.74	0.10	-0.23	-1.50	-2.30
-1.40	-1.97	0.10	-0.16	-1.40	-1.61
-1.30	-2.13	0.10	-0.09	-1.30	-0.92
-1.20	-2.22	0.10	-0.03	-1.20	-0.26
-1.10	-2.25	0.10	0.04	-1.10	0.39
-1.00	-2.21	0.10	0.10	-1.00	1.01
-0.90	-2.11	0.10	0.16	-0.90	1.60
-0.80	-1.95	0.10	0.22	-0.80	2.16
-0.70	-1.73	0.10	0.27	-0.70	2.68
-0.60	-1.46	0.10	0.32	-0.60	3.16
-0.50	-1.15	0.10	0.36	-0.50	3.60
-0.40	-0.79	0.10	0.40	-0.40	3.99
-0.30	-0.39	0.10	0.43	-0.30	4.34



Example 3.3.12: secant lines

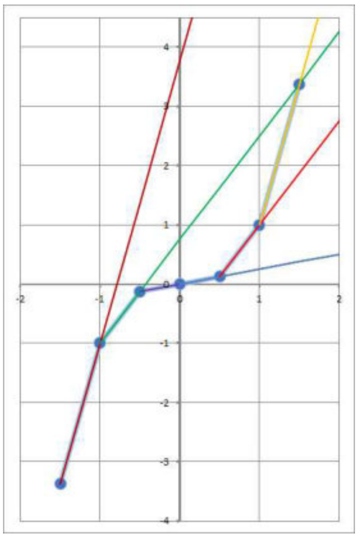
Let's consider $f(x)=x^2$. Below, we use spreadsheets to sample the function at these 5 points, draw lines through every two consecutive ones, and compute the slope of each:



These are a couple of computations:

- The slope from $(0,0)$ to $(1,1)$ is $\dfrac{1-0}{1-0}=1$.
- The slope from $(1,1)$ to $(2,4)$ is $\dfrac{4-1}{2-1}=3$.

Now $f(x)=x^3$:

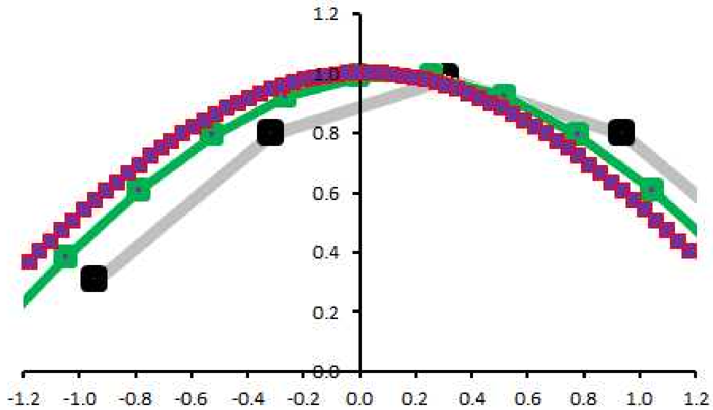


A couple of computations:

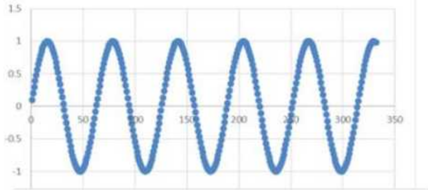
- The slope from $(0,0)$ to $(1,1)$ is $\dfrac{1-0}{1-0}=1$.
- The slope from $(1,1)$ to $(2,8)$ is $\dfrac{8-1}{2-1}=7$.

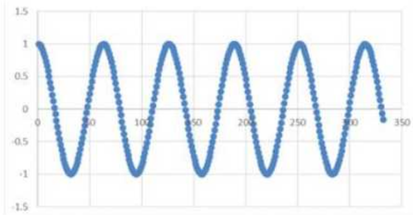
Example 3.3.13: denser sampling

Let's apply the formula to $f(x)=\sin x$. Different values of the increment of x produce different values of the difference quotient. However, decreasing this number shows a pattern of *convergence*:



In the first graph below, we sample this function every $\Delta x = .1$:





The second graph is the difference quotient. It looks like $y = \cos x$, especially if we shift it a little to the right.

Based on these definitions, we can use the language of motion in a more precise manner:

Definition 3.3.14: velocity

When a function or a sequence is called "location" or "position", its difference quotient is called the *velocity*.

Definition 3.3.15: acceleration

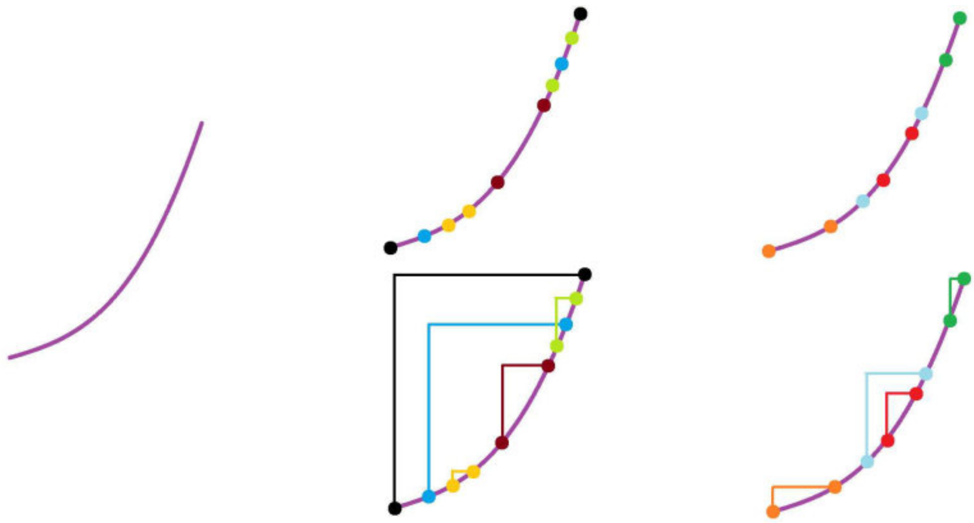
When a function or a sequence is called "velocity", its difference quotient is called the *acceleration*.

3.4. *The limit of the difference quotient: the derivative*

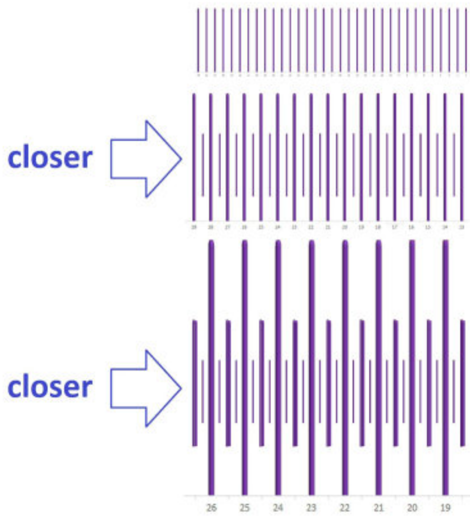
In general, suppose $y = f(x)$ is a function defined on an interval $[a, b]$. Then, we have a function that we can freely *sample*, i.e., to find its value $y = f(x)$ for any choice of x. Then the computation of the rates of change presented in the last section is available for every pair of values of x! Indeed, it is the slope of the line from $(x_0, f(x_0))$ to $(x_1, f(x_1))$ on the graph of the function:

$$\boxed{\frac{f(x_1) - f(x_0)}{x_1 - x_0}}$$

There are infinitely many such computations for a single function:



They aren't unstructured though; the sample points are arranged in partitions of the interval:



In order to make sense of all the rates of change of a function, we will be making these partitions denser and denser.

Example 3.4.1: denser sampling

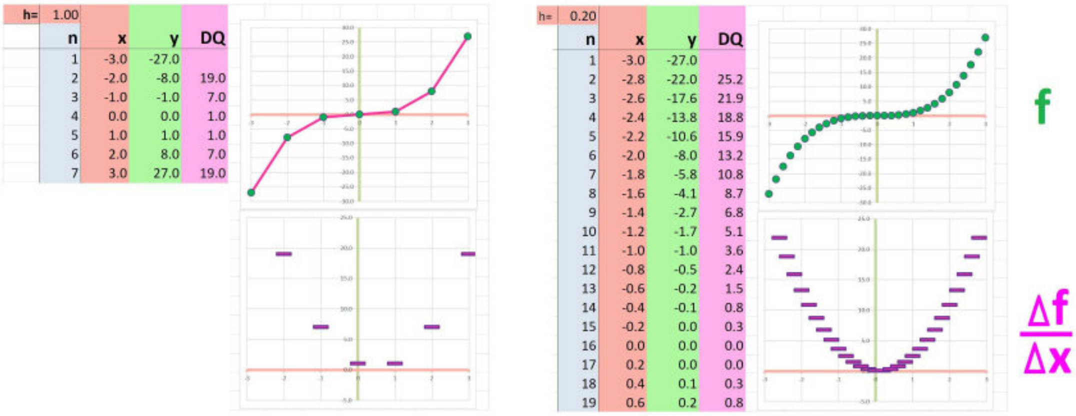
We let the spreadsheet compute the difference quotients for $f(x) = x^2$ (left) and plot the slopes in a separate chart below:

h=	1.00		
n	x	y	DQ
1	-3.0	9.0	
2	-2.0	4.0	-5.0
3	-1.0	1.0	-3.0
4	0.0	0.0	-1.0
5	1.0	1.0	1.0
6	2.0	4.0	3.0
7	3.0	9.0	5.0

h=	0.20		
n	x	y	DQ
0	-3.0	9.0	
1	-2.8	7.8	-5.8
2	-2.6	6.8	-5.4
3	-2.4	5.8	-5.0
4	-2.2	4.8	-4.6
5	-2.0	4.0	-4.2
6	-1.8	3.2	-3.8
7	-1.6	2.6	-3.4
8	-1.4	2.0	-3.0
9	-1.2	1.4	-2.6
10	-1.0	1.0	-2.2
11	-0.8	0.6	-1.8
12	-0.6	0.4	-1.4
13	-0.4	0.2	-1.0
14	-0.2	0.0	-0.6
15	0.0	0.0	-0.2
16	0.2	0.0	0.2
17	0.4	0.2	0.6
18	0.6	0.4	1.0

We then repeat the procedure for a denser pattern of points (right).

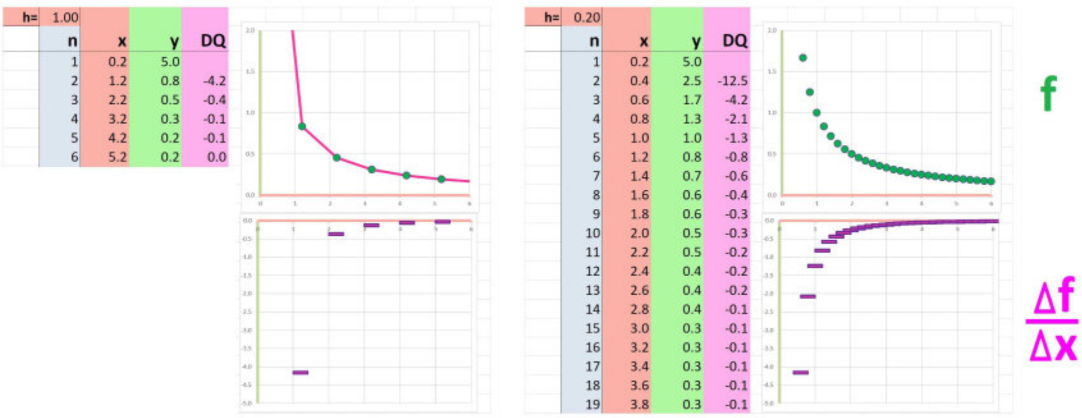
Now $f(x) = x^3$. We compute the difference quotient using a spreadsheet for the two values of the increment $h = 1$ and $h = .2$:



h=	1.00		
n	x	y	DQ
1	-3.0	-27.0	
2	-2.0	-8.0	19.0
3	-1.0	-1.0	7.0
4	0.0	0.0	1.0
5	1.0	1.0	1.0
6	2.0	8.0	7.0
7	3.0	27.0	19.0

h=	0.20		
n	x	y	DQ
1	-3.0	-27.0	
2	-2.8	-22.0	25.2
3	-2.6	-17.6	21.9
4	-2.4	-13.8	18.8
5	-2.2	-10.6	15.9
6	-2.0	-8.0	13.2
7	-1.8	-5.8	10.8
8	-1.6	-4.1	8.7
9	-1.4	-2.7	6.8
10	-1.2	-1.7	5.1
11	-1.0	-1.0	3.6
12	-0.8	-0.5	2.4
13	-0.6	-0.2	1.5
14	-0.4	-0.1	0.8
15	-0.2	0.0	0.3
16	0.0	0.0	0.0
17	0.2	0.0	0.0
18	0.4	0.1	0.3
19	0.6	0.2	0.8

This is the difference quotient $\frac{\Delta f}{\Delta x}$, i.e., the sampled slopes, of $f(x) = 1/x$ (bottom row):



h=	1.00		
n	x	y	DQ
1	0.2	5.0	
2	1.2	0.8	-4.2
3	2.2	0.5	-0.4
4	3.2	0.3	-0.1
5	4.2	0.2	-0.1
6	5.2	0.2	0.0

h=	0.20		
n	x	y	DQ
1	0.2	5.0	
2	0.4	2.5	-12.5
3	0.6	1.7	-4.2
4	0.8	1.3	-2.1
5	1.0	1.0	-1.3
6	1.2	0.8	-0.8
7	1.4	0.7	-0.6
8	1.6	0.6	-0.4
9	1.8	0.6	-0.3
10	2.0	0.5	-0.3
11	2.2	0.5	-0.2
12	2.4	0.4	-0.2
13	2.6	0.4	-0.2
14	2.8	0.4	-0.1
15	3.0	0.3	-0.1
16	3.2	0.3	-0.1
17	3.4	0.3	-0.1
18	3.6	0.3	-0.1
19	3.8	0.3	-0.1

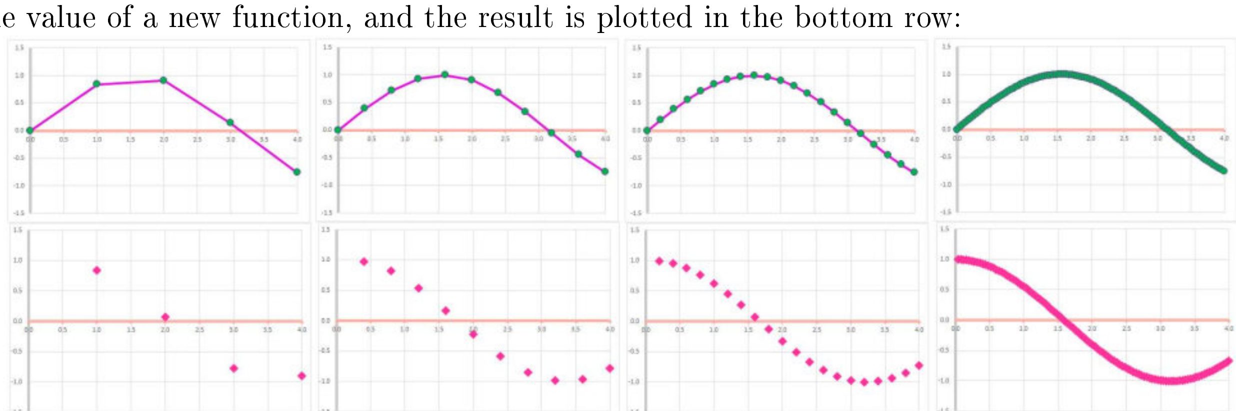
Exercise 3.4.2

Do the new graphs look familiar?

Example 3.4.3: $\sin x$

Let's take another look at $f(x) = \sin x$. A function is given below, and so is the interval $[a, b]$.

For each $n = 2, 3, 4, ...$, the increment is found, $\Delta x = (b - a)/n$, and we have n segments in our partition of $[a, b]$. On each of the segments, the difference quotient is computed, the value is recorded as the value of a new function, and the result is plotted in the bottom row:



We face a *sequence of functions*. Should we think of its limit?

Our sampling of the function is getting denser and denser. In the meantime, the points that make up the graph of the difference quotient are getting closer and closer together. What is at the end of this process? A new function defined on the whole interval!

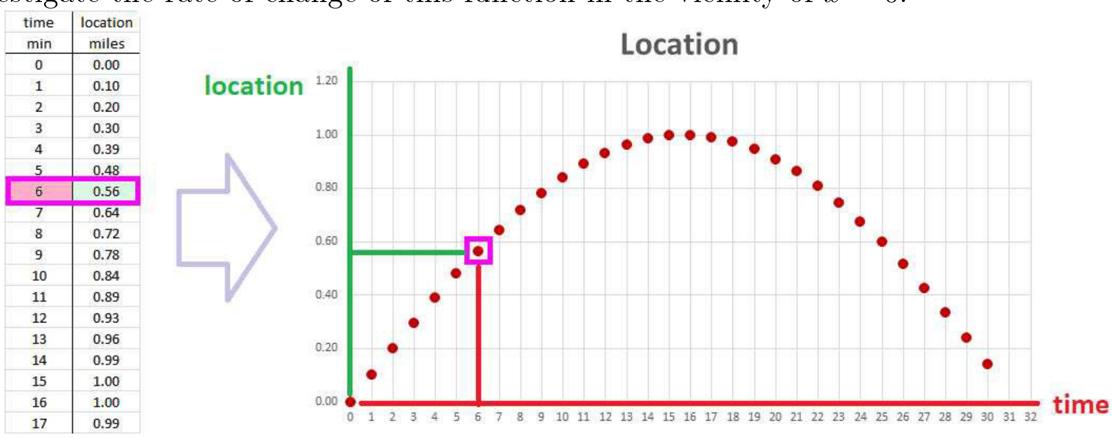
Even though the difference quotients above seem to produce recognizable new functions, it is a challenge to capture the whole function at once. Initially, our approach will be the following:

▶ The difference quotients are evaluated in the vicinity of a *single point*.

Once again, we have a function that we can freely *sample*, i.e., to find its value $y = f(x)$ for any choice of x, and we would like to make sense of the rate of change of this function *in the vicinity* of a particular value of x.

Example 3.4.4: decrease Δx

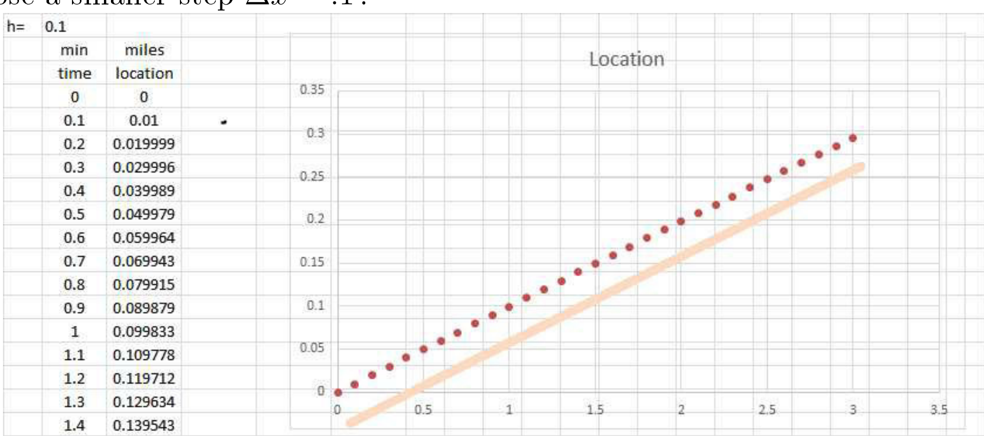
Let's investigate the rate of change of this function in the vicinity of $x = 0$:



time	location
min	miles
0	0.00
1	0.10
2	0.20
3	0.30
4	0.39
5	0.48
6	0.56
7	0.64
8	0.72
9	0.78
10	0.84
11	0.89
12	0.93
13	0.96
14	0.99
15	1.00
16	1.00
17	0.99

These values of x have come from a partition with the original node $a = x_0 = 0$ and the increment $\Delta x = 1$. The rate of change, i.e., difference quotient of the function for this partition, varies.

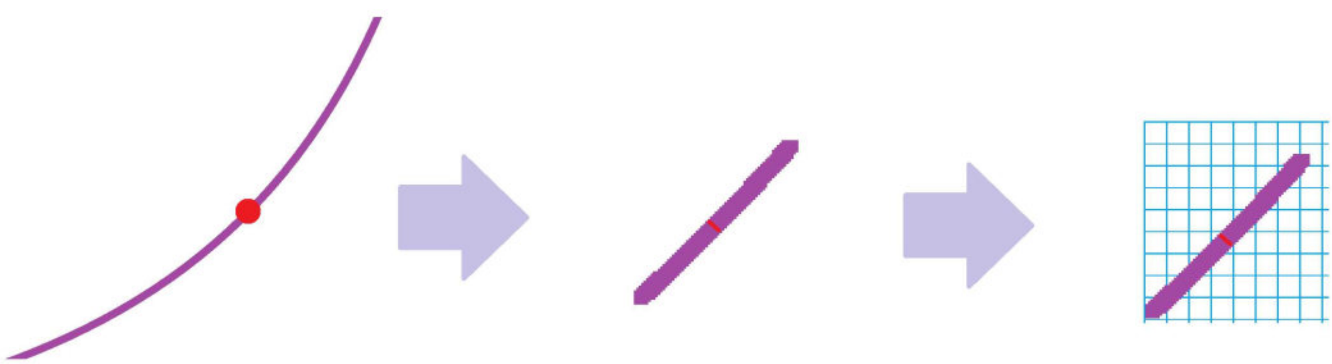
Let's now choose a smaller step $\Delta x = .1$:



h=	0.1
min	miles
time	location
0	0
0.1	0.01
0.2	0.019999
0.3	0.029996
0.4	0.039989
0.5	0.049979
0.6	0.059964
0.7	0.069943
0.8	0.079915
0.9	0.089879
1	0.099833
1.1	0.109778
1.2	0.119712
1.3	0.129634
1.4	0.139543

We can see an almost straight line! In fact, the difference quotient is almost constant (the slope appears to be .1). We were using *two-node* partitions here, $\{a,\ a + \Delta x\}$.

The *derivative* is the slope of the tangent line. One can estimate this number from a picture by blowing up a small piece surrounding the point and then placing it on a grid:



But what *is* the tangent line?

Example 3.4.5: sequence of secants

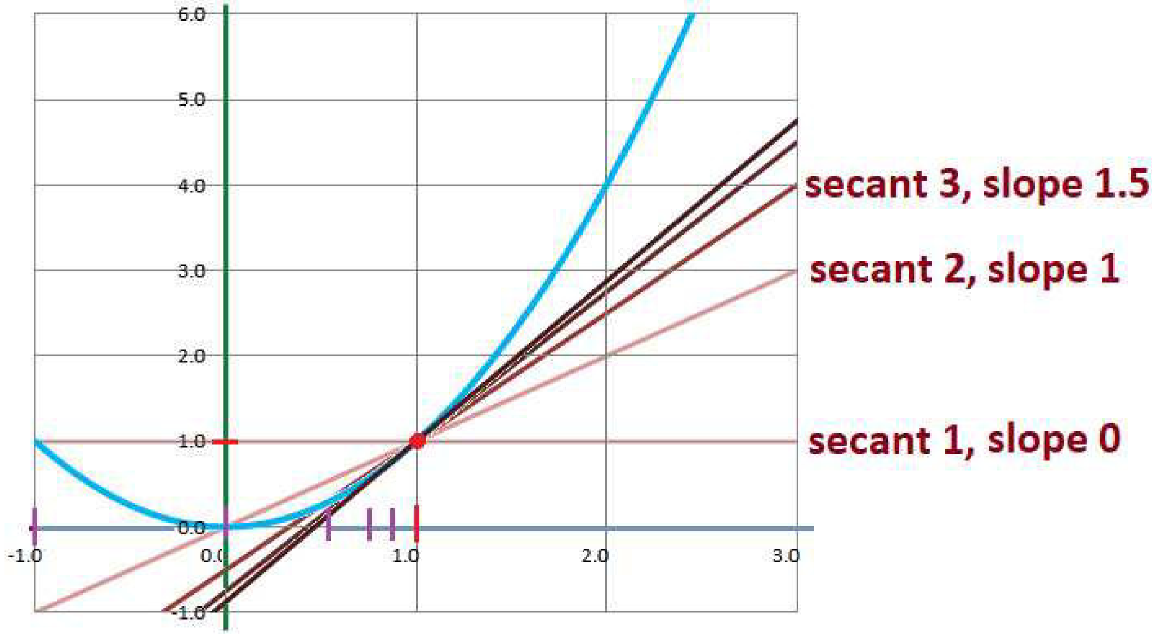
Let's consider $y = x^2$ at $x = 1$. Five inputs are chosen closer and closer to 1 from the left:

$$x = -1,\ 0,\ .5,\ .75,\ .875\,.$$

Then we find corresponding points on the graph:

$$(x, y) = (-1, 1),\ (0, 0),\ (.5, .25),\ (.75, .5625),\ (.875, 0.765625)\,.$$

Then a line is drawn through $(1, 1)$ and each of these points.



These lines are supposed to converge onto the tangent line. However, it is much easier to concentrate on the slopes only: instead of a sequence of lines, we have a sequence of numbers! For each of these values of x, we have a line drawn through $(1, 1)$ and (x, x^2). Therefore, the slopes are:

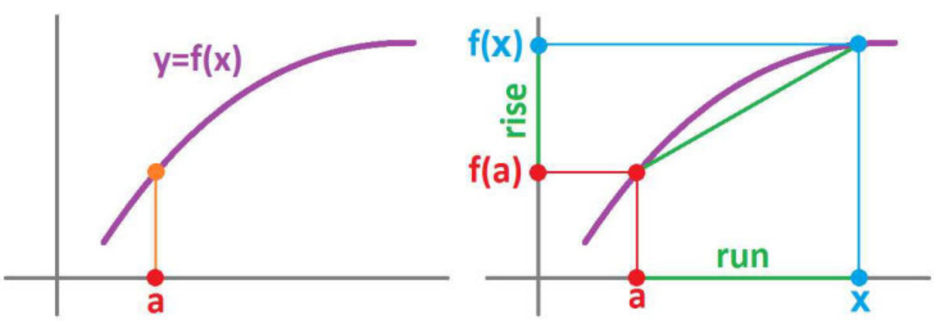
$$\frac{x^2 - 1}{x - 1}\,.$$

From our computations in the last chapter, we know that the limit of this expression as x is approaching 1 is 2. That's the slope of the tangent line!

So, instead of trying to understand what the tangent line is, we simply find its slope.

We define its slope as the limit of slopes of the cords, called the *secant lines*, i.e., lines determined by two points on the graph. For each $x \neq a$, the slope of the line through $(a, f(a))$ and $(x, f(x))$ is:

$$\text{slope } = \frac{\text{rise}}{\text{run}} = \frac{f(x) - f(a)}{x - a}\,.$$



The difference quotient of a function f at a is defined for each $x \neq a$ to be

$$\frac{f(x) - f(a)}{x - a}\,.$$

Then the slope of the tangent line is the limit of these slopes. As we move x toward a, the secant lines *turn* and approach the tangent line, provided the limit exists. The algebraic representation of this idea is below:

Definition 3.4.6: derivative at a point

The *derivative of f at $x = a$* is defined to be the limit of the difference quotients of f at $x = a$ as the increment Δx is approaching 0, denoted as follows:

$$\boxed{\frac{df}{dx}(a) = \lim_{x\to a} \frac{f(x) - f(a)}{x - a}}$$

Warning!

The derivative is *not* a fraction; this is just a notation.

Below is the breakdown of the notation:

Derivative

$$\text{not a fraction:}\qquad \begin{array}{ccc} & \text{old function} & \\ & \downarrow & \\ & \dfrac{df}{dx} & (a) \\ \nearrow & & \uparrow \\ \text{new function} & & \text{input variable} \end{array}$$

Alternatively written, the formula reveals the reason for the notation:

$$\boxed{\frac{df}{dx}(a) = \lim_{\Delta x\to 0} \frac{\Delta f}{\Delta x}}$$

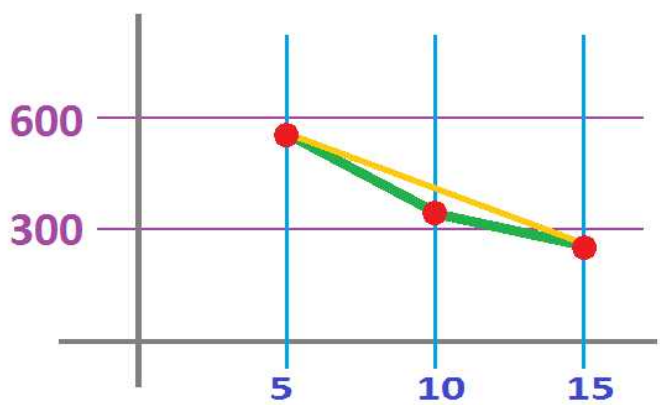
Example 3.4.7: table of values

Let's estimate the derivative of a function given by numbers only:

x	5	10	15
$y = f(x)$	554	344	250

What is $\dfrac{df}{dx}(10)$?

We use the difference quotient, i.e., the slope of a secant line. There are several choices:



We can use the slope of either of the two adjacent secant lines:

- Slope of the 1st segment $= \dfrac{344-554}{5} = -42\,.$
- Slope of the 2nd segment $= \dfrac{250-344}{5} = -18.8\,.$

We can also use the average of the two.

$$\text{Average slope } = \frac{-42+(-18.8)}{2} = \frac{-60.8}{2} = -30.4\,.$$

The last option is to use the two segments as a single interval:

$$\text{Slope } = \frac{250-554}{10} = -30.4\,.$$

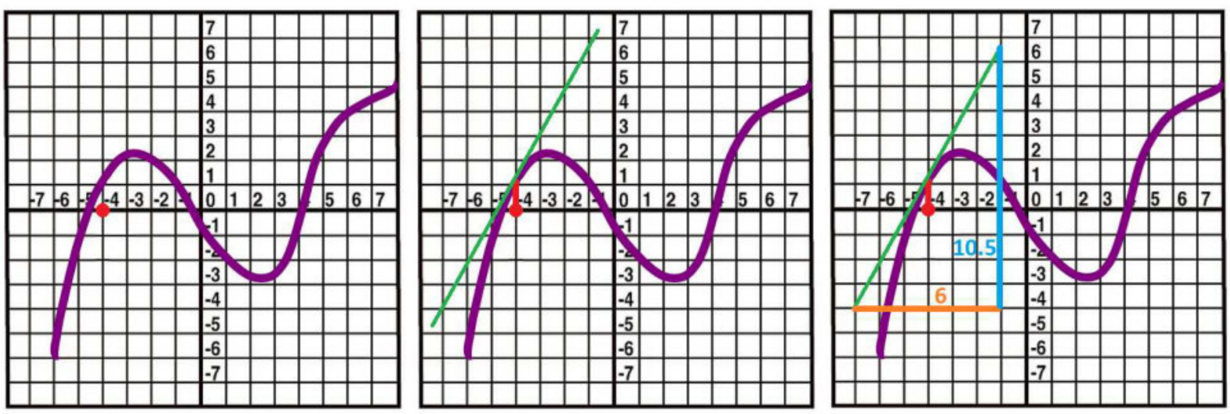
It's the same number!

Exercise 3.4.8

Is this a coincidence?

Example 3.4.9: derivative from graph

One can estimate the derivative from the graph by using our ability to plot tangent lines. Let's find the derivative at the red point:



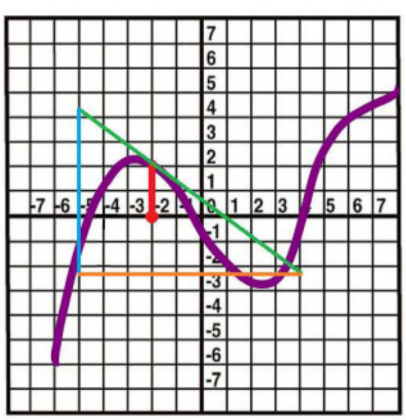
These are the steps:

1. Plot the tangent line (green).
2. Using the grid, build a right triangle with the segment of the tangent line as its hypotenuse.
3. Compute the slope:

$$\text{slope } = \frac{\text{rise}}{\text{run}} \approx \frac{10.5}{6} = 1.75\,.$$

To better estimate the slope, draw the triangle as large as possible.

Find the derivative at the point given below:



To find the slope, we need the rise. The height of the triangle is 6.5. However, the only way to justify

using the word "rise" while we are going *down* is to give it a negative value:

$$\text{rise} \ = -6.5\,.$$

Then the slope and the derivative are also negative:

$$\frac{df}{dx}(3) = \frac{-6.5}{9} \approx -7.2\,.$$

Exercise 3.4.10

Find the derivative at the rest of the integer points.

Example 3.4.11: graph from derivatives

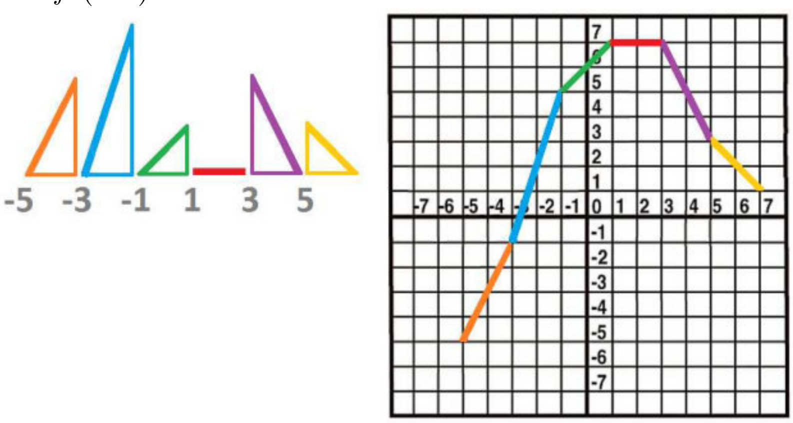
Now we go in the opposite direction: Plot the graph based on numerical data. Suppose only these values of the derivative are known:

$$\frac{df}{dx}(-5) = 2,\ \frac{df}{dx}(-3) = 3,\ \frac{df}{dx}(-1) = 1,\ \frac{df}{dx}(1) = 0,\ \frac{df}{dx}(3) = -2,\ \frac{df}{dx}(5) = -1\,.$$

Each is the slope of the respective tangent line and, therefore, the rise of this line whenever the run is equal to 1. Instead, we choose the run to be the distance between the points, i.e., 2. Therefore, the values of the rise are:

x :	-5	-3	-1	1	3	5
rise:	4	6	2	0	-4	-2

Then we have six triangles and their hypotenuses are meant to make up rough approximation of the segment of the graph of f (left):



In light of the discussions in the last chapter, we want f to be *continuous*! That is why we attach each piece to the end of the last one (right).

Exercise 3.4.12

Try to plot the graphs of other functions with these values of the derivative.

Exercise 3.4.13

Plot the graph of f based on these values of the derivative:

$$\frac{df}{dx}(-5) = -1,\ \frac{df}{dx}(-4) = 1,\ \frac{df}{dx}(-3) = 0,\ \frac{df}{dx}(-2) = 2,\ \frac{df}{dx}(-1) = -2,\ \frac{df}{dx}(0) = 3\,.$$

Make up your own numbers and repeat.

Next, we go beyond approximations.

3.5. The derivative is the instantaneous rate of change

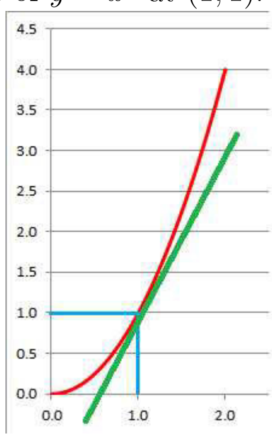
The main idea of this chapter is as follows:

- ▶ The derivative is the limit of the difference quotient.

We will now use our skills with limits (Chapter 2) to compute derivatives.

Example 3.5.1: x^2

Find the derivative and the tangent line of $y = x^2$ at $(1, 1)$:



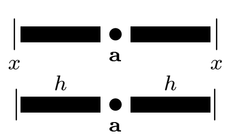
Use $f(x) = x^2$, $a = 1$ in the definition:

$$\begin{aligned}
\frac{df}{dx}(a) &= \lim_{x\to a} \frac{f(x) - f(a)}{x - a} \\
&= \lim_{x\to 1} \frac{x^2 - 1^2}{x - 1} && \to \frac{0}{0}? \quad \boxed{\textbf{DEAD END}} \\
&= \lim_{x\to 1} \frac{(x-1)(x+1)}{x - 1} \\
&= \lim_{x\to 1} (x + 1) \\
&= 1 + 1 \\
&= 2\,. && \text{That's the slope!}
\end{aligned}$$

Then the point-slope form of the line is:

$$y - 1 = 2(x - 1)\,.$$

In Chapter 2, we saw an alternative formula for the limit: Instead of concentrating on how x is approaching a, we look at the (signed) distance $h = x - a$ between them. The two approaches are illustrated below:



Then for any function g we have:

$$\lim_{x\to a} g(x) = \lim_{h\to 0} g(a + h)\,.$$

We apply this idea to the limit of the derivative.

Let's consider the formula for the difference quotient. Here, instead of concentrating on how x is approaching a, we look at the (signed) distance $h = x - a$ between them. Then $h \to 0$. We substitute $h = x - a$ into the

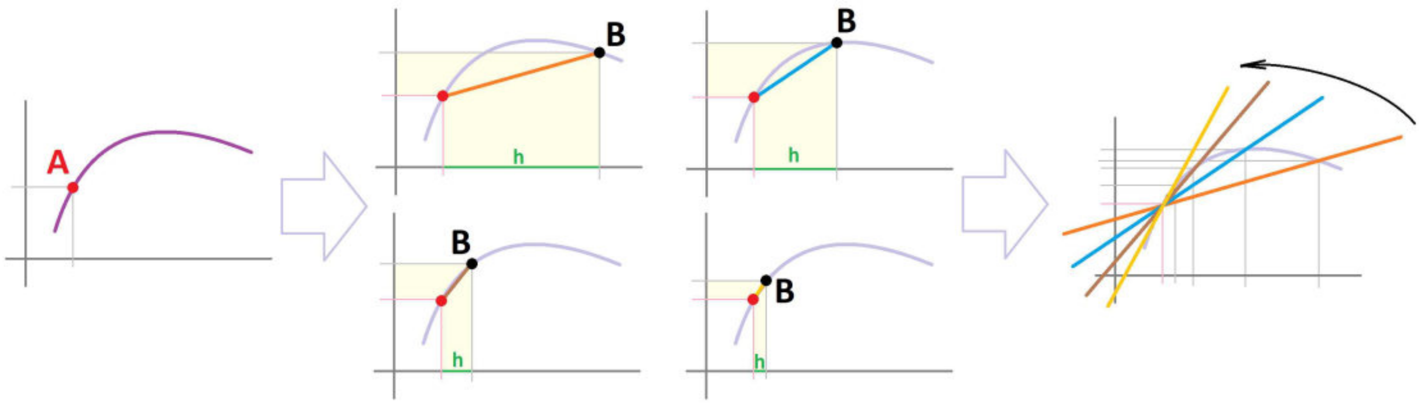
definition (top) and obtain an alternative formula (bottom) after the substitution $x = a + h$:

$$\frac{df}{dx}(a) = \lim_{x\to a} \frac{f(x) - f(a)}{\underbrace{x - a}_{h=x-a}}$$

$$\frac{df}{dx}(a) = \lim_{h\to 0} \frac{f(a+h) - f(a)}{h}$$

Thus, what this substitution has accomplished is a *change of variables* in the limit.

The geometry of the limit remains the same:



Here, the secant segments are getting shorter and shorter so that it is hard to tell what is going on. We extend them into lines. As you can see, these lines keep turning: 2nd is closer to the tangent than 1st, etc. The computation is entirely about *angles*!

Warning!

The derivative is *not* the limit of f, $\lim\limits_{x\to a} f(x)$, but a limit of a new function made (derived) from f.

Example 3.5.2: x^2

This substitution makes computations easier sometimes. Let's consider the last example again:

$$f(x) = x^2 \,.$$

Compute the difference quotient and its limit:

$$\begin{aligned}
\frac{df}{dx}(1) &= \lim_{h\to 0} \frac{f(1+h) - f(1)}{h} \\
&= \lim_{h\to 0} \frac{(1+h)^2 - 1^2}{h} && \to \frac{0}{0}? \quad \boxed{\textbf{DEAD END}} \\
&= \lim_{h\to 0} \frac{1^2 + 2h + h - 1}{h} \\
&= \lim_{h\to 0} \frac{2h + h^2}{h} \\
&= \lim_{h\to 0} (2 + h) && \text{The cancellation.} \\
&= 2 + 0 = 2 \,.
\end{aligned}$$

Exercise 3.5.3

Find $\frac{df}{dx}(-1)$ and $\frac{df}{dx}(0)$ for $f(x) = x^2$.

Exercise 3.5.4

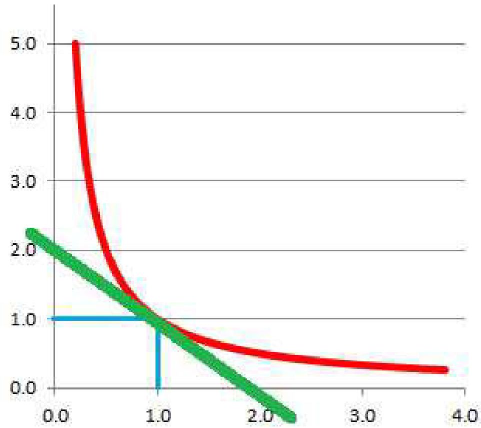
Find $\frac{df}{dx}(1)$ for $f(x) = x^2 + 1$.

Exercise 3.5.5

Find $\frac{df}{dx}(1)$ for $f(x) = x^3$.

Example 3.5.6: $1/x$

Find $\frac{df}{dx}(1)$ for $f(x) = \frac{1}{x}$.



$$\begin{aligned}
\frac{df}{dx} &= \lim_{h\to 0} \frac{f(1+h) - f(1)}{h} \\
&= \lim_{h\to 0} \frac{\frac{1}{1+h} - \frac{1}{1}}{h} && \to \frac{0}{0}? \qquad \boxed{\textbf{DEAD END}} \\
&= \lim_{h\to 0} \frac{\frac{1}{1+h} - \frac{1+h}{1+h}}{h} \\
&= \lim_{h\to 0} \left(\frac{1-(1-h)}{1+h} \cdot \frac{1}{h} \right) \\
&= \lim_{h\to 0} \left(\frac{-h}{1+h} \cdot \frac{1}{h} \right) \\
&= \lim_{h\to 0} \left(-\frac{1}{1+h} \right) && \text{The cancellation.} \\
&= -\frac{1}{1+0} = -1\,.
\end{aligned}$$

At the end, recognizing that this is a rational function continuous on its domain, we simply plug in the value.

Exercise 3.5.7

Find $\frac{df}{dx}(1)$ for $f(x) = 1/x^2$.

Note that every time we compute the derivative as a (special kind of) limit, we might go through the same stages:

▶ Quotient Rule (error!) ... leading to ... an indeterminate expression ... leading to ... algebra ... leading to ... the cancellation ... leading to ... substitution

We should just ignore what we know to be a dead-end and go straight to algebra!

Sometimes algebra is not enough though.

Example 3.5.8: trigonometry

First $f(x) = \sin x$. Then

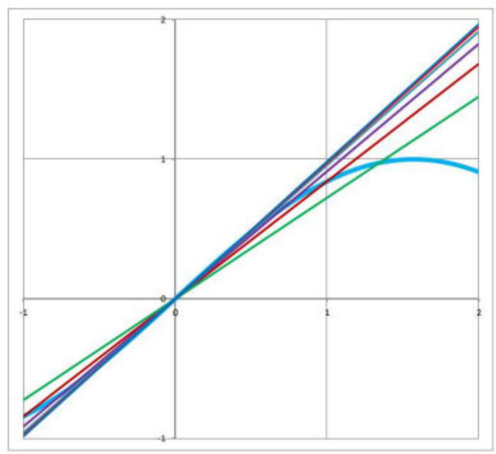
$$\begin{aligned} \left.\frac{\Delta f}{\Delta x}\right|_{x=0} &= \frac{\sin h - \sin 0}{h} \\ &= \frac{\sin h}{h} && \text{No cancellation.} \\ &\to 1 && \text{as } h \to 0\,. \end{aligned}$$

The last step is simply the following famous limit from Chapter 2:

$$\lim_{x\to 0} \frac{\sin x}{x} = 1\,.$$

The result is confirmed with a spreadsheet:

	n=		1	2	3	4	5	6	7	8	9	10	11	12	13
	x=	0.00	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10	0.09	0.08	0.08
	y=	0.00	0.84	0.48	0.33	0.25	0.20	0.17	0.14	0.12	0.11	0.10	0.09	0.08	0.08
	slope=		0.84	0.72	0.91	0.96	0.97	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00
h	x	f	y												
0.1	-1.00	-0.84	-0.84	-0.72	-0.91	-0.96	-0.97	-0.98	-0.99	-0.99	-0.99	-0.99	-1.00	-1.00	-1.00
	-0.90	-0.78	-0.76	-0.65	-0.82	-0.86	-0.88	-0.88	-0.89	-0.89	-0.89	-0.89	-0.90	-0.90	-0.90
	-0.80	-0.72	-0.67	-0.58	-0.73	-0.77	-0.78	-0.79	-0.79	-0.79	-0.79	-0.80	-0.80	-0.80	-0.80
	-0.70	-0.64	-0.59	-0.51	-0.64	-0.67	-0.68	-0.69	-0.69	-0.69	-0.70	-0.70	-0.70	-0.70	-0.70
	-0.60	-0.56	-0.50	-0.43	-0.55	-0.57	-0.58	-0.59	-0.59	-0.59	-0.60	-0.60	-0.60	-0.60	-0.60
	-0.50	-0.48	-0.42	-0.36	-0.46	-0.48	-0.49	-0.49	-0.49	-0.50	-0.50	-0.50	-0.50	-0.50	-0.50
	-0.40	-0.39	-0.34	-0.29	-0.37	-0.38	-0.39	-0.39	-0.40	-0.40	-0.40	-0.40	-0.40	-0.40	-0.40
	-0.30	-0.30	-0.25	-0.22	-0.27	-0.29	-0.29	-0.29	-0.30	-0.30	-0.30	-0.30	-0.30	-0.30	-0.30
	-0.20	-0.20	-0.17	-0.14	-0.18	-0.19	-0.19	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20	-0.20
	-0.10	-0.10	-0.08	-0.07	-0.09	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10	-0.10
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	0.10	0.10	0.08	0.07	0.09	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10	0.10
	0.20	0.20	0.17	0.14	0.18	0.19	0.19	0.20	0.20	0.20	0.20	0.20	0.20	0.20	0.20
	0.30	0.30	0.25	0.22	0.27	0.29	0.29	0.29	0.30	0.30	0.30	0.30	0.30	0.30	0.30
	0.40	0.39	0.34	0.29	0.37	0.38	0.39	0.39	0.40	0.40	0.40	0.40	0.40	0.40	0.40
	0.50	0.48	0.42	0.36	0.46	0.48	0.49	0.49	0.49	0.50	0.50	0.50	0.50	0.50	0.50
	0.60	0.56	0.50	0.43	0.55	0.57	0.58	0.59	0.59	0.59	0.60	0.60	0.60	0.60	0.60
	0.70	0.64	0.59	0.51	0.64	0.67	0.68	0.69	0.69	0.69	0.70	0.70	0.70	0.70	0.70
	0.80	0.72	0.67	0.58	0.73	0.77	0.78	0.79	0.79	0.79	0.79	0.80	0.80	0.80	0.80
	0.90	0.78	0.76	0.65	0.82	0.86	0.88	0.88	0.89	0.89	0.89	0.89	0.90	0.90	0.90
	1.00	0.84	0.84	0.72	0.91	0.96	0.97	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00



In other words, *the graph of* $y = \sin x$ *crosses the y-axis at* 45 *degrees.*

Exercise 3.5.9

Find $\frac{df}{dx}(\pi)$ for $f(x) = \sin x$.

Exercise 3.5.10

Find $\frac{df}{dx}(\pi)$ for $f(x) = \sin x + x$.

Example 3.5.11: trigonometry

Second, $f(x) = \cos x$. Then

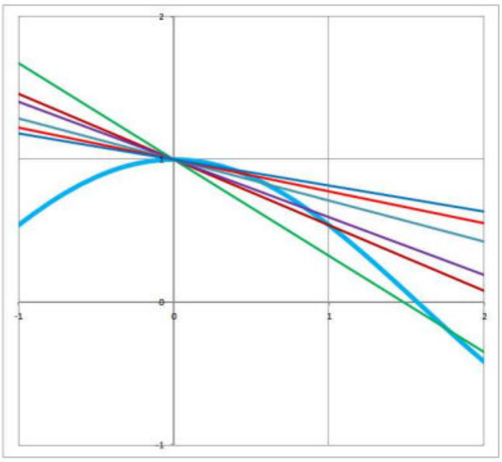
$$\begin{aligned} \left.\frac{\Delta f}{\Delta x}\right|_{x=0} &= \frac{\cos h - \cos 0}{h} \\ &= \frac{\cos h - 1}{h} && \text{No cancellation.} \\ &\to 0 && \text{as } h \to 0\,. \end{aligned}$$

The last step is simply the other famous limit from Chapter 2:

$$\lim_{x\to 0} \frac{1-\cos x}{x} = 0\,.$$

The result is confirmed with a spreadsheet:

	n=		1	2	3	4	5	6	7	8	9	10	11	12	13
	x=	0.00	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10	0.09	0.08	0.08
	y=	1.00	0.54	0.88	0.94	0.97	0.98	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
	slope=		-0.46	-0.67	-0.40	-0.29	-0.22	-0.18	-0.15	-0.13	-0.12	-0.11	-0.10	-0.09	-0.08
h	x	f	y												
0.1	-1.00	0.54	1.46	1.67	1.40	1.29	1.22	1.18	1.15	1.13	1.12	1.11	1.10	1.09	1.08
	-0.90	0.62	1.41	1.61	1.36	1.26	1.20	1.16	1.14	1.12	1.11	1.09	1.09	1.08	1.07
	-0.80	0.70	1.37	1.54	1.32	1.23	1.18	1.15	1.12	1.11	1.09	1.08	1.08	1.07	1.06
	-0.70	0.76	1.32	1.47	1.28	1.20	1.16	1.13	1.11	1.09	1.08	1.07	1.07	1.06	1.06
	-0.60	0.83	1.28	1.40	1.24	1.17	1.13	1.11	1.09	1.08	1.07	1.06	1.06	1.05	1.05
	-0.50	0.88	1.23	1.34	1.20	1.14	1.11	1.09	1.08	1.07	1.06	1.05	1.05	1.04	1.04
	-0.40	0.92	1.18	1.27	1.16	1.11	1.09	1.07	1.06	1.05	1.05	1.04	1.04	1.03	1.03
	-0.30	0.96	1.14	1.20	1.12	1.09	1.07	1.05	1.05	1.04	1.04	1.03	1.03	1.03	1.02
	-0.20	0.98	1.09	1.13	1.08	1.06	1.04	1.04	1.03	1.03	1.02	1.02	1.02	1.02	1.02
	-0.10	1.00	1.05	1.07	1.04	1.03	1.02	1.02	1.02	1.01	1.01	1.01	1.01	1.01	1.01
	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	1.00	0.95	0.93	0.96	0.97	0.98	0.98	0.98	0.99	0.99	0.99	0.99	0.99	0.99
	0.20	0.98	0.91	0.87	0.92	0.94	0.96	0.96	0.97	0.97	0.98	0.98	0.98	0.98	0.98
	0.30	0.96	0.86	0.80	0.88	0.91	0.93	0.95	0.95	0.96	0.96	0.97	0.97	0.97	0.98
	0.40	0.92	0.82	0.73	0.84	0.89	0.91	0.93	0.94	0.95	0.95	0.96	0.96	0.97	0.97
	0.50	0.88	0.77	0.66	0.80	0.86	0.89	0.91	0.92	0.93	0.94	0.95	0.95	0.96	0.96
	0.60	0.83	0.72	0.60	0.76	0.83	0.87	0.89	0.91	0.92	0.93	0.94	0.94	0.95	0.95
	0.70	0.76	0.68	0.53	0.72	0.80	0.84	0.87	0.89	0.91	0.92	0.93	0.93	0.94	0.94
	0.80	0.70	0.63	0.46	0.68	0.77	0.82	0.85	0.88	0.89	0.91	0.92	0.92	0.93	0.94
	0.90	0.62	0.59	0.39	0.64	0.74	0.80	0.84	0.86	0.88	0.89	0.91	0.91	0.92	0.93
	1.00	0.54	0.54	0.33	0.60	0.71	0.78	0.82	0.85	0.87	0.88	0.89	0.90	0.91	0.92



In other words, *the graph of $y=\cos x$ crosses the y-axis horizontally.*

Exercise 3.5.12

Find $\frac{df}{dx}(\pi)$ for $f(x)=\cos x$.

Exercise 3.5.13

Find $\frac{df}{dx}(0)$ for $f(x)=\sin x+\cos x$.

Example 3.5.14: exponential

Let $f(x)=e^x$. Then

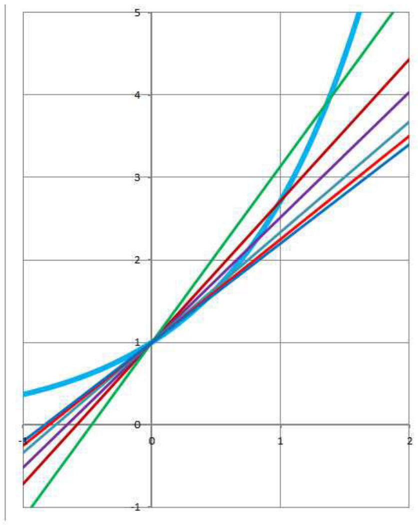
$$\begin{aligned} \frac{\Delta f}{\Delta x}\bigg|_{x=0} &= \frac{e^h-e^0}{h} \\ &= \frac{e^h-1}{h} && \text{No cancellation.} \\ &\to 1 && \text{as } h\to 0\,. \end{aligned}$$

The last step is simply the following famous limit from Chapter 2:

$$\lim_{x\to 0} \frac{e^x-1}{x} = 1\,.$$

The result is confirmed with a spreadsheet:

	n=		1	2	3	4	5	6	7	8	9	10	11	12	13
	x=	0.00	1.00	0.50	0.33	0.25	0.20	0.17	0.14	0.13	0.11	0.10	0.09	0.08	0.08
	y=	1.00	2.72	1.65	1.40	1.28	1.22	1.18	1.15	1.13	1.12	1.11	1.10	1.09	1.08
	slope=		1.72	2.14	1.52	1.34	1.25	1.20	1.17	1.14	1.13	1.11	1.10	1.09	1.08
h	x	f	y												
0.1	-1.00	0.37	-0.72	-1.14	-0.52	-0.34	-0.25	-0.20	-0.17	-0.14	-0.13	-0.11	-0.10	-0.09	-0.08
	-0.90	0.41	-0.55	-0.93	-0.37	-0.21	-0.13	-0.08	-0.05	-0.03	-0.01	0.00	0.01	0.02	0.02
	-0.80	0.45	-0.37	-0.71	-0.21	-0.07	0.00	0.04	0.07	0.09	0.10	0.11	0.12	0.13	0.13
	-0.70	0.50	-0.20	-0.50	-0.06	0.06	0.12	0.16	0.18	0.20	0.21	0.22	0.23	0.24	0.24
	-0.60	0.55	-0.03	-0.28	0.09	0.20	0.25	0.28	0.30	0.31	0.32	0.33	0.34	0.35	0.35
	-0.50	0.61	0.14	-0.07	0.24	0.33	0.37	0.40	0.42	0.43	0.44	0.44	0.45	0.45	0.46
	-0.40	0.67	0.31	0.14	0.39	0.46	0.50	0.52	0.53	0.54	0.55	0.56	0.56	0.56	0.57
	-0.30	0.74	0.48	0.36	0.54	0.60	0.62	0.64	0.65	0.66	0.66	0.67	0.67	0.67	0.67
	-0.20	0.82	0.66	0.57	0.70	0.73	0.75	0.76	0.77	0.77	0.77	0.78	0.78	0.78	0.78
	-0.10	0.90	0.83	0.79	0.85	0.87	0.87	0.88	0.88	0.89	0.89	0.89	0.89	0.89	0.89
	0.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.10	1.11	1.17	1.21	1.15	1.13	1.13	1.12	1.12	1.11	1.11	1.11	1.11	1.11	1.11
	0.20	1.22	1.34	1.43	1.30	1.27	1.25	1.24	1.23	1.23	1.23	1.22	1.22	1.22	1.22
	0.30	1.35	1.52	1.64	1.46	1.40	1.38	1.36	1.35	1.34	1.34	1.33	1.33	1.33	1.33
	0.40	1.49	1.69	1.86	1.61	1.54	1.50	1.48	1.47	1.46	1.45	1.44	1.44	1.44	1.43
	0.50	1.65	1.86	2.07	1.76	1.67	1.63	1.60	1.58	1.57	1.56	1.56	1.55	1.55	1.54
	0.60	1.82	2.03	2.28	1.91	1.80	1.75	1.72	1.70	1.69	1.68	1.67	1.66	1.65	1.65
	0.70	2.01	2.20	2.50	2.06	1.94	1.88	1.84	1.82	1.80	1.79	1.78	1.77	1.76	1.76
	0.80	2.23	2.37	2.71	2.21	2.07	2.00	1.96	1.93	1.91	1.90	1.89	1.88	1.87	1.87
	0.90	2.46	2.55	2.93	2.37	2.21	2.13	2.08	2.05	2.03	2.01	2.00	1.99	1.98	1.98
	1.00	2.72	2.72	3.14	2.52	2.34	2.25	2.20	2.17	2.14	2.13	2.11	2.10	2.09	2.08



In other words, *the graph of $y = e^x$ crosses the y-axis at 45 degrees.* It follows that the natural base exponential function cuts all the rest in two halves: steep and shallow.

Exercise 3.5.15

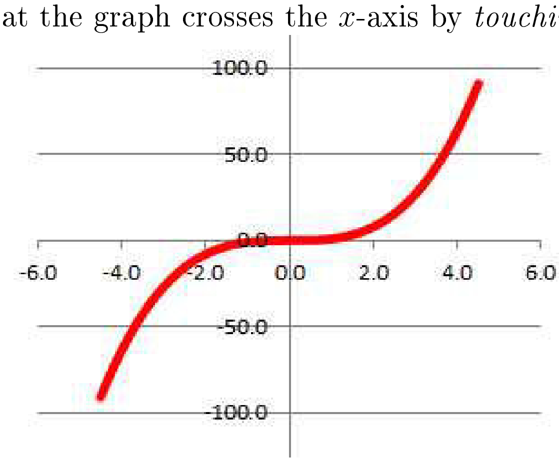
Find $\dfrac{df}{dx}(1)$ for $f(x) = e^x$.

Example 3.5.16: x^3

Compute the derivative of $f(x) = x^3$ at $x = a$. Let's take all difference quotients at once:

$$\frac{x^3 - a^3}{x - a} = x^2 + xa + a^2 \to 3a^2 \quad \text{as} \quad x \to a\,.$$

In particular, we conclude that the graph crosses the x-axis by *touching* it:



Let's check:

- If we try $a = -4$, the derivative is 48.
- If we try $a = 0$, the derivative is 0.
- If we try $a = 4$, the derivative is 48.

The results match the picture.

3.6. The existence of the derivative: differentiability

We have computed simplified formulas for

1. the difference quotients, and then
2. the derivatives.

The former precedes the latter and takes care of all the algebra. The latter requires applications of the methods of computing limits presented in Chapter 1.

The process and the methods of finding the explicit formulas for functions given by explicit formulas is called *differentiation*.

Warning!

The word "differentiate" has nothing to do with "distinguish" or "tell apart".

Some limits don't exist. Then, as a limit, the derivative,

$$\frac{df}{dx}(a) = \lim_{x\to a} \frac{f(x)-f(a)}{x-a}\,,$$

might not exist either.

Example 3.6.1: undefined

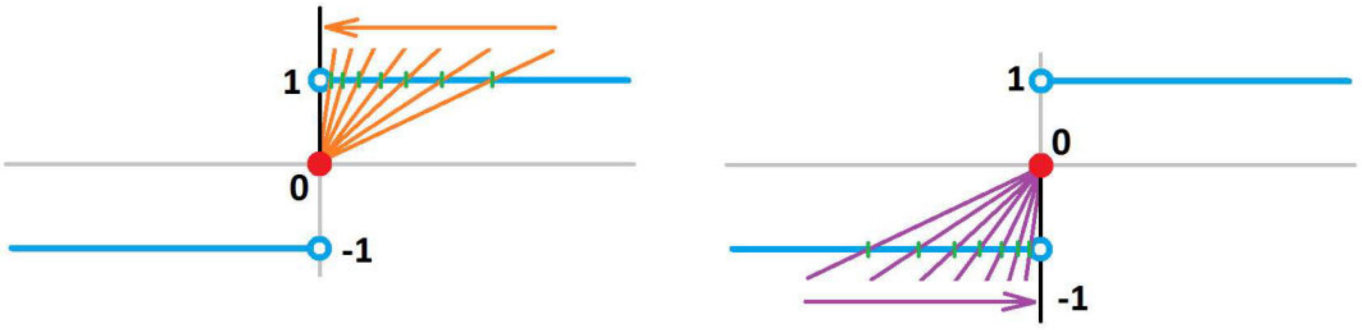
To begin with, if a function is undefined at a, the numerator is undefined and neither is the derivative at a. For example, $\frac{df}{dx}(0)$ doesn't exist for $f(x)=1/x$.

Example 3.6.2: $\operatorname{sign}(x)$

Let's consider a discontinuous function,

$$y=f(x)=\operatorname{sign}(x)\,.$$

We see below that as x is approaching 0, the secant lines become more and more steep and approach a vertical one:



This means that slopes $\to +\infty$. So, their limit $\frac{df}{dx}(a)$ is infinite.

This is how this fact effects the algebra:

$$\begin{aligned}\frac{df}{dx}(0) &= \lim_{x\to 0}\frac{\operatorname{sign}(x)-\operatorname{sign}(0)}{x-0}\\ &= \lim_{x\to 0}\frac{\operatorname{sign}(x)}{x}\\ &= \lim_{x\to 0}\frac{1}{|x|}\\ &= \infty\,.\end{aligned}$$

This limit does not exist. The derivative doesn't exist.

Even if we suppose that the function *is* defined at $x = a$ and the secant lines are defined, it is possible that as $h \to 0$ they do not tend toward any particular line.

In our motion metaphor, the function represents the location and the derivative is the velocity. If the location changes instantly (as in the last example), the velocity must be infinite:

$$\text{velocity} = \frac{\text{distance} \neq 0}{\text{time} = 0}\,.$$

When this limit does exist, what does it tell us? It tells us that there is a tangent line at that location. What does that tell us? It tells us that there is no *break* in the graph!

Theorem 3.6.3: Derivative Needs Continuity

If the derivative of f exists at a, then f is continuous at a.

Proof.

Let's consider the *rise*, for $x \neq a$, and rewrite it with the help of the difference quotient as follows:

$$f(x) - f(a) = \frac{f(x) - f(a)}{x - a} \cdot (x - a)\,.$$

It's the same function. Let's take the limit of this function as $x \to a$:

$$\begin{array}{ccccc} \underbrace{\dfrac{f(x) - f(a)}{x - a}} & \cdot & \underbrace{(x - a)} & = & \underbrace{f(x) - f(a)} \\ \downarrow & & \downarrow & & \downarrow \\ f'(a) & \cdot & 0 & \Longrightarrow & 0 \end{array}$$

So, the limit of the first factor exists and is equal to $f'(a)$, and the limit of the second factor exists and is equal to 0. Therefore, the *Product Rule* applies, and we conclude that the limit of the right-hand side is also zero:

$$f(x) - f(a) = \frac{f(x) - f(a)}{x - a} \cdot (x - a) \to f'(a) \cdot 0 = 0\,.$$

The conclusion,

$$\lim_{x\to a}(f(x) - f(a)) = 0\,,$$

is rewritten according to the *Sum Rule* as follows:

$$\lim_{x\to a} f(x) - \lim_{x\to a} f(a) = 0\,.$$

Now, since the limit of a constant function is the constant, $\lim_{x\to a} f(a) = f(a)$, we have:

$$\lim_{x\to a} f(x) = f(a)\,.$$

The identity means that f is continuous at a.

Indeed, on many occasions, we zoomed in on the graph of a function to discover that it looks like a straight line. The idea suggests that those functions were continuous! We will apply this term to these functions:

Definition 3.6.4: differentiable function

When the limit of the difference quotient at a,

$$\lim_{x\to a} \frac{f(x) - f(a)}{x - a}\,,$$

exists, we say that f is *differentiable at a*.

Thus, we say:

- ▶ Every differentiable function is also continuous.

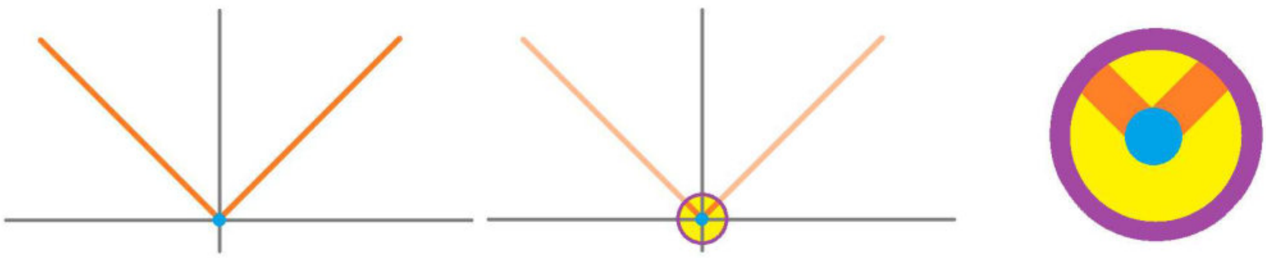
Exercise 3.6.5

Present the last statement as an implication. State the converse too.

The converse isn't true.

Example 3.6.6: absolute value function

If we zoom in on the graph of the absolute value function around $(0,0)$, it won't become a straight line:



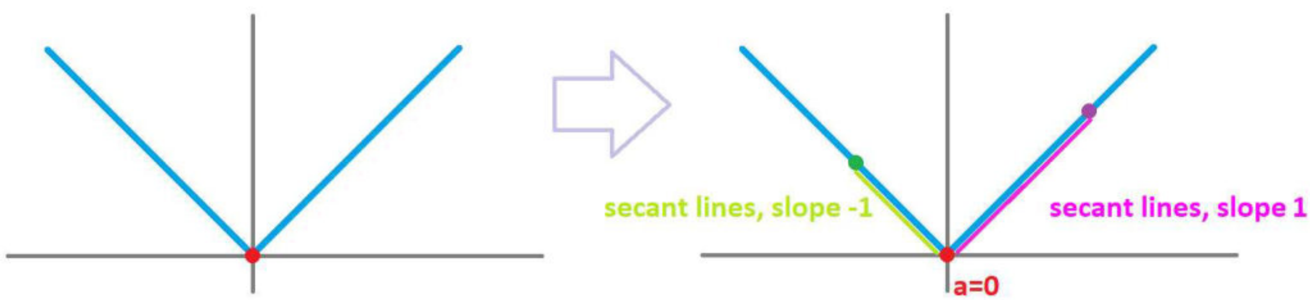
The corner of the V won't disappear even after multiple tries. Let's confirm this algebraically by trying to compute the derivative of $f(x)=|x|$ at $x=0$. We have:

$$\begin{aligned}\frac{df}{dx}(0) &= \lim_{h\to 0}\frac{|0+h|-|0|}{h}\\ &= \lim_{h\to 0}\frac{|h|}{h}\,.\end{aligned}$$

We've seen this limit before! Consider these one-sided limits:

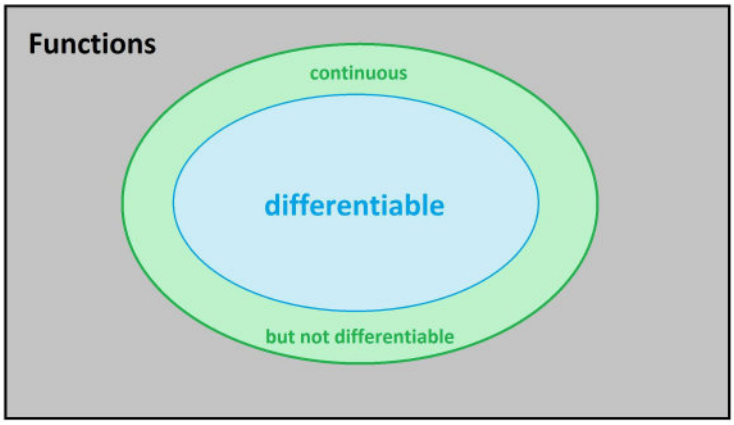
$$\begin{aligned}\lim_{h\to 0^-}\frac{|h|}{h} &= \lim_{h\to 0^-}\frac{-h}{h} &= -1\\ \lim_{h\to 0^+}\frac{|h|}{h} &= \lim_{h\to 0^+}\frac{h}{h} &= 1\end{aligned}$$

They are not equal, so the limit does not exist. We see this fact below:



The secant lines simply follow the line of the graph itself.

We can visualize these two main classes of functions:



Exercise 3.6.7

Place within the diagram: $|x|$, $\operatorname{sign} x$, x^3.

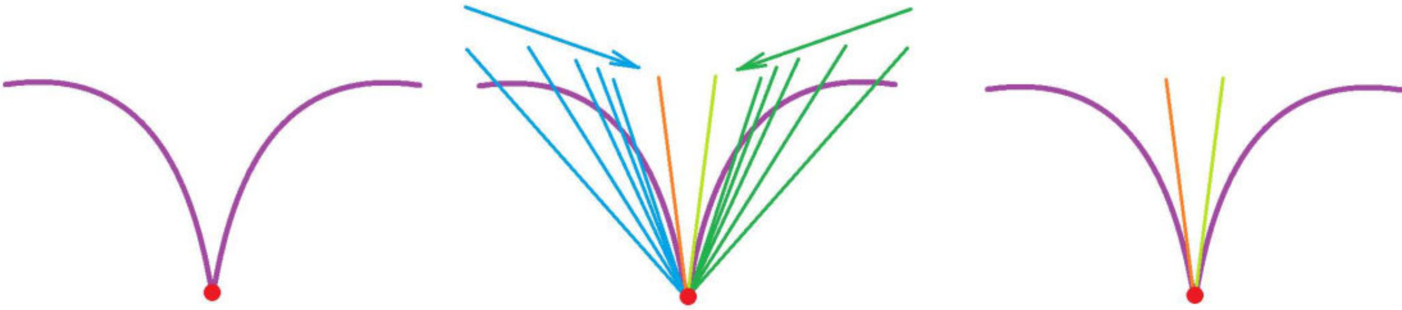
Compare:

- Continuous: there is no break or gap.
- Differentiable: there is no corner or cusp.

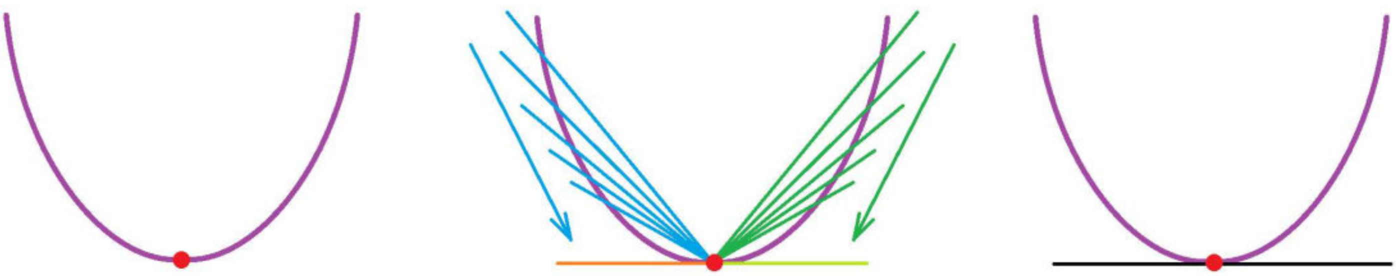
As we just saw, when there are two candidates to be a tangent, there is no tangent!

Let's try to use the geometric definition of the derivative – via secant lines. In fact, we will use secant *rays*. As we approach a separately from the left and right, they turn, and the end result is one of these two:

- Two rays. Then f is not differentiable:



- One line. Then f is differentiable:



Algebraically, we verify that the two one-sided limits are equal to each other:

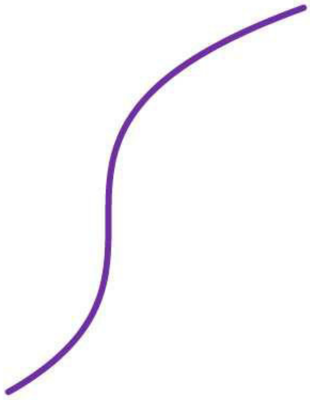
$$\lim_{h\to 0^-} = \lim_{h\to 0^+}$$

Exercise 3.6.8

What if the end result is *one* ray?

Example 3.6.9: smooth graphs

There are functions the graph of which look *smooth* (no cusps) and yet they are not differentiable:

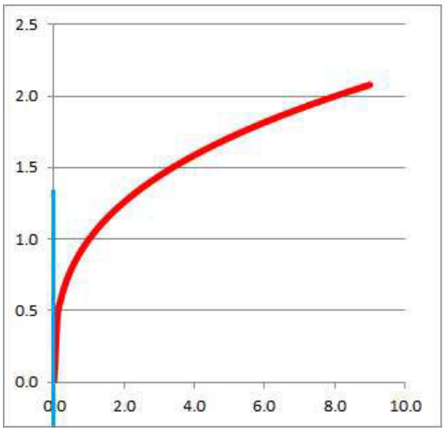


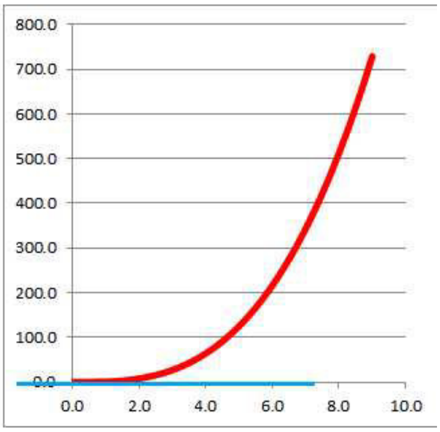
Indeed, if we zoom in on the origin, the line will appear vertical!

Consider a specific function:

$$f(x) = \sqrt[3]{x}$$

at $x = 0$ (left):





The limit is infinite even without a computation:

$$\frac{df}{dx}(0) = \infty\,.$$

How do we know? This function is the inverse of $x = y^3$ (right) and the derivative of the latter at 0 is 0; therefore, its tangent line is horizontal at 0. Then, the tangent line of $y = \sqrt[3]{x}$ at 0 is vertical!

Example 3.6.10: quadratic

Compute $\dfrac{df}{dx}(2)$ from the definition for

$$f(x) = -x^2 - x\,.$$

Definition:

$$\frac{df}{dx}(2) = \lim_{h\to 0} \frac{f(2+h) - f(2)}{h}\,.$$

To compute the difference quotient, we need to substitute twice. In $f(x) = -x^2 - x$, we replace x with $2 + h$ and then we replace x with 2:

$$f(2+h) = -(2+h)^2 - (2+h),\;\; f(2) = -2^2 - 2\,.$$

Now, we substitute into the definition:

$$\begin{aligned}\frac{df}{dx}(2) &= \lim_{h\to 0}\frac{[-(2-h)^2-(2-h)]-[-2^2-2]}{h}\\ &= \lim_{h\to 0}\frac{-4-4h-h^2-2-h+4+2}{h}\\ &= \lim_{h\to 0}\frac{-5h-h^2}{h}\\ &= \lim_{h\to 0}(-5-h)\\ &= -5-0\\ &= 5\,.\end{aligned}$$

Example 3.6.11: income tax

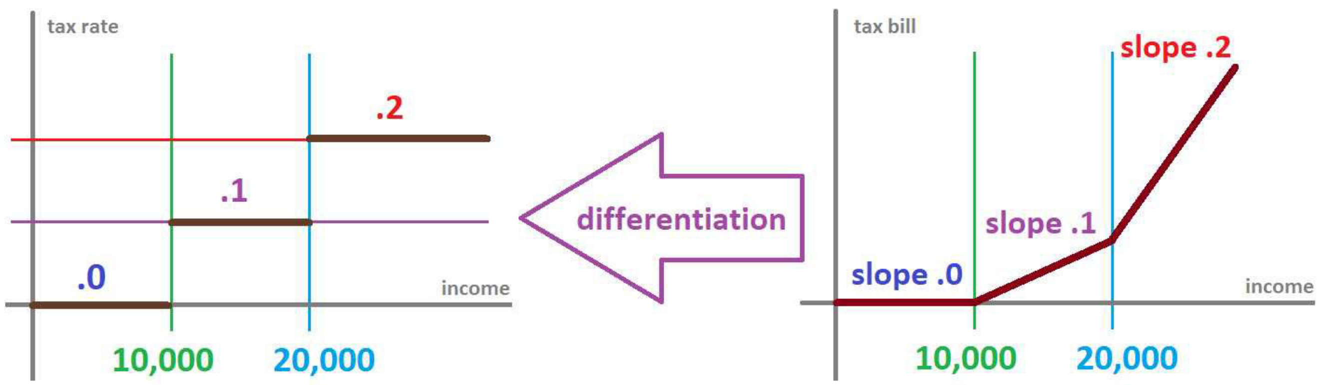
Recall the hypothetical tax code from Chapter 1. If x is the income, then the marginal tax *rate* is computed by the formula:

$$f(x)=\begin{cases}0 & \text{if } x\le 10000\,,\\ .10 & \text{if } 10000<x\le 20000\,,\\ .20 & \text{if } 20000<x\,.\end{cases}$$

The tax bill as a function of income is as follows:

$$g(x)=\begin{cases}0 & \text{if } x\le 10000\,,\\ .10\cdot(x-10000) & \text{if } 10000<x\le 20000\,,\\ .10\cdot(x-10000)+.20\cdot(x-20000) & \text{if } 20000<x\,.\end{cases}$$

Let's plot both:



Looking at the graphs, we see that the slopes of the latter are the values of the former! In other words, the former is the derivative of the latter:

$$\frac{dg}{dx}(x)=f(x)\,,$$

except for the values of x where g isn't differentiable: $x=10,000$ and $x=20,000$.

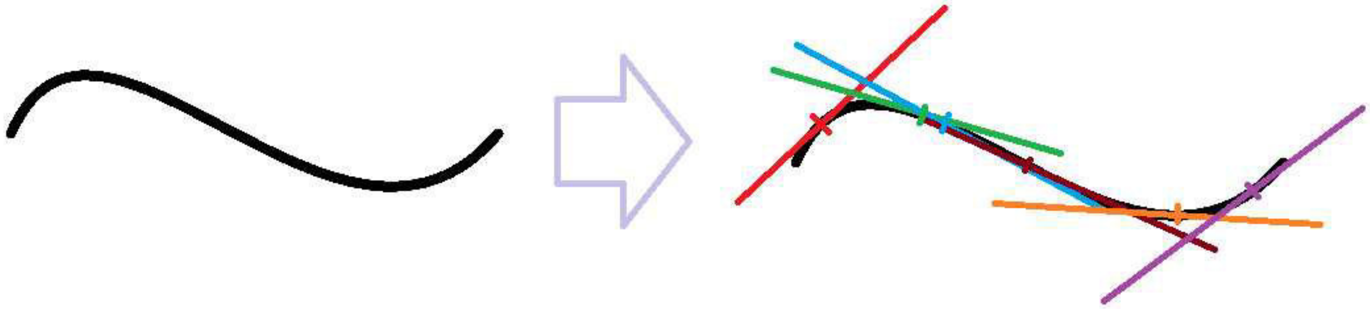
The points where the function is differentiable form the *domain of the derivative*.

3.7. The derivative as a function

The difference quotient is a function, defined at the nodes of the partition. What about the derivative?

What if we construct the secant and the tangent lines to the graph of a function *at all points* at the same

time? It looks complicated:



The solution is algebraic: We find, or attempt to find, the derivative – as the limit of the difference quotient – at every point a of the domain of f. These a's are the *inputs* of the new function, and the found values of the derivative $\frac{df}{dx}(a)$ are the *outputs*. The result is the *derivative function*! Its domain consists of all points where f is differentiable.

This is an alternative, more compact, notation for the derivative:

Derivative

$$f'(x) = \frac{df}{dx}(x)$$

It reads "f prime".

Since x is the same independent variable for both, it makes sense to use the that name for the input of the derivative:

$$y = f(x) \longrightarrow z = f'(x)\,.$$

Thus, a new function has been *derived* from the old, and so are the names: the name of the new function makes a reference to the function it came from. Let's deconstruct the notation:

Lagrange notation and Leibniz notation for derivative

name of function	name of input	name of function	name of input
f	x	f	x
f'	x	$\frac{df}{dx}$	x

$$f'(x) = \frac{df}{dx}(x)$$

$$f' = \frac{df}{dx}$$

The choice between the two is a matter of convenience.

Warning!

The name of the *output* of the derivative doesn't match that of the function.

In order to build a theory around the new concept, we start with the simplest situations.

Consider this *obvious* statement about motion:

- "If I am standing still, my speed is zero."

If a function $y = f(x)$ represents the position, we can restate this mathematically. We follow what we know about the differences of sequences (seen in Volume 1, Chapter 1PC-1) with the following:

Theorem 3.7.1: Difference of Constant Function

If a function defined at the nodes of a partition of interval $[a, b]$ is constant over the nodes of $[a, b]$, then the function has a zero difference for all secondary nodes in the partition.

In other words, we have:

$$f(x) = \text{ constant } \implies \Delta f(c) = 0\,.$$

We just divide by Δx to prove the following:

Theorem 3.7.2: Difference Quotient of Constant Function

If a function defined at the nodes of a partition of interval $[a, b]$ is constant over the nodes of $[a, b]$, then the function has a zero difference quotient for all secondary nodes in the partition.

In other words, we have:

$$f(x) = \text{ constant } \implies \frac{\Delta f}{\Delta x}(c) = 0\,.$$

Taking the limit $\Delta x \to 0$ proves the following:

Theorem 3.7.3: Derivative of Constant Function

If a function is constant on an open interval I, then its derivative is zero for all x in I.

In other words, we have:

$$f(x) = \text{ constant } \implies \frac{df}{dx}(x) = 0\,.$$

In summary:

$$f(x) = \text{ constant } \implies \Delta f(c) = 0 \implies \frac{\Delta f}{\Delta x}(c) = 0 \implies \frac{df}{dx}(x) = 0\,.$$

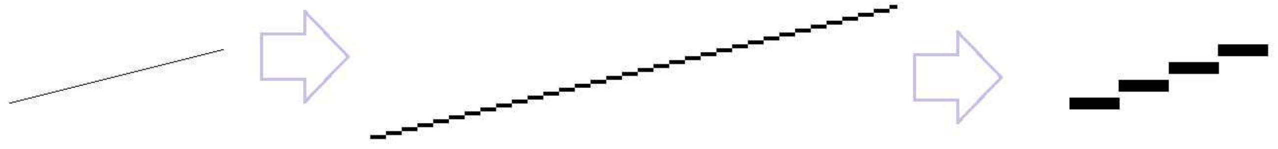
Exercise 3.7.4

Provide details of the proof.

The converse is also true but its proof is postponed until Chapter 5.

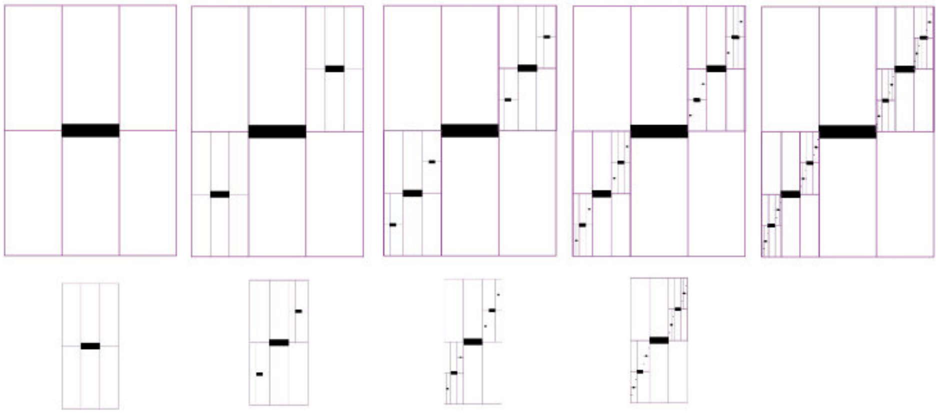
Example 3.7.5: piecewise constant functions

Is it possible to have a function with zero derivative but not constant? Zooming in reveals gaps:

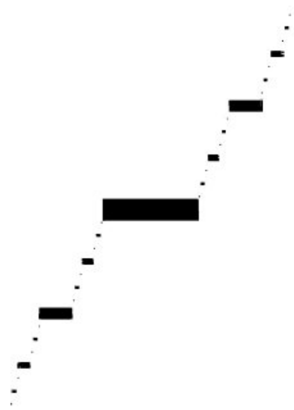


Example 3.7.6: Cantor's staircase

To face the challenge, let's build a function that gets close to this impossibility. We cut the rectangle in half vertically and in three horizontally, then place a horizontal segment in the middle:



The we do the same with the two rectangles located diagonally from the segment, and so on. When we remove the scaffolding, we have the following graph:



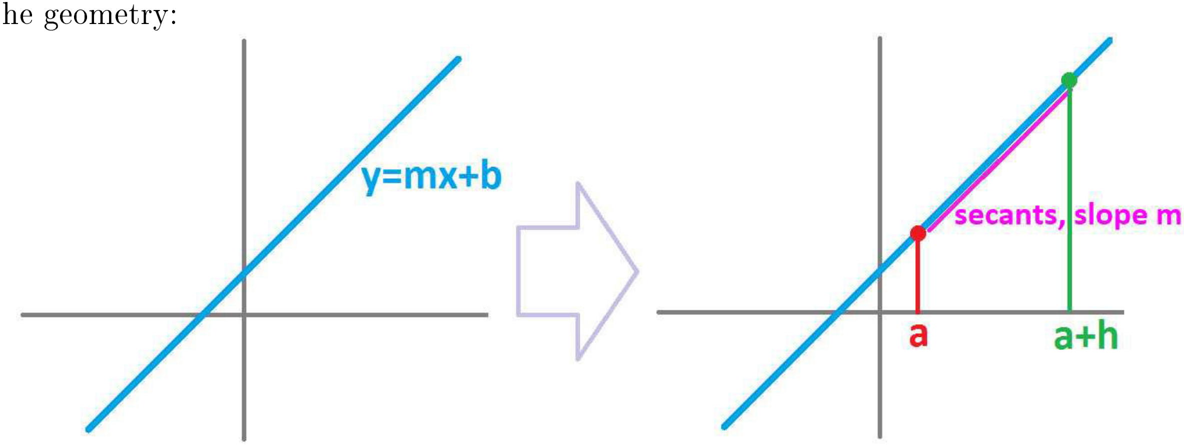
The function is known to be continuous and increasing even though made entirely of horizontal pieces.

Example 3.7.7: general linear function

Let's compute the derivative at $x = a$ of a linear function

$$f(x) = mx + b\,.$$

First, the geometry:

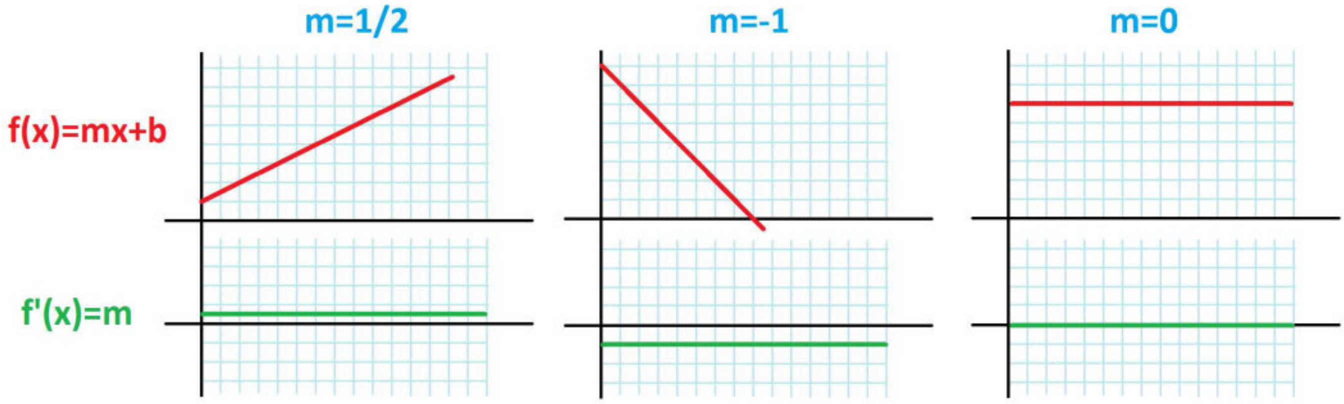


Every secant line connects our point of interest, $(a, f(a))$, to another point on the graph, $(a+h, f(a+h))$, where $h = \Delta x$ is the increment of the independent variable. Then, the secant line lies entirely within the straight line that is the graph of f. It then has the same slope. Therefore, the derivative must be m!

Now, algebra: The difference quotient is independent of location. Indeed:

$$\begin{aligned}
\frac{\Delta f}{\Delta x}(a) &= \frac{f(a+h)-f(a)}{h} && \text{Substitute } x=a+h \text{ into the formula of } f.\\
&= \frac{[m(a+h)+b]-[ma+b]}{h} && \text{Simplify.}\\
&= \frac{mh}{h} && \text{Then divide.}\\
&= m && \Longrightarrow\\
\frac{df}{dx}(a) &= \lim_{h\to 0}\frac{\Delta f}{\Delta x}(a) = m\,.
\end{aligned}$$

The result is a *number*. However, since is it independent of the chosen a, we treat it as a *function*, a constant function:



The example proves the first half of the following statement:

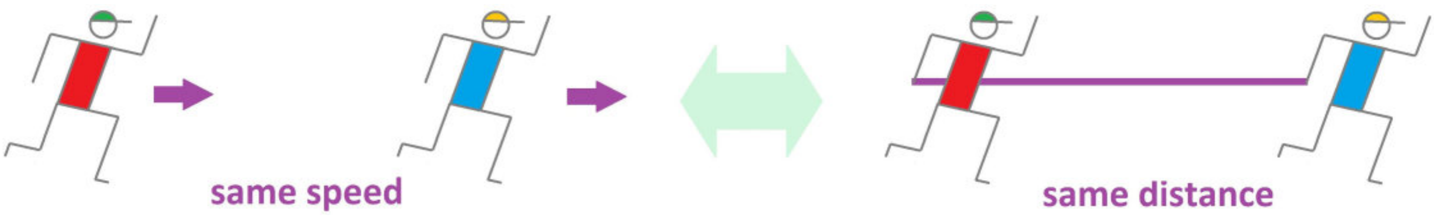
▶ The derivative of a linear function is constant and, conversely, a constant function can only be the derivative of a linear function.

The second half will remain a conjecture for now.

To continue with our theory, suppose this time that there are *two* runners; we have a slightly less obvious fact about motion:

▶ "If the distance between two runners isn't changing, then they run with the same speed."

It's as if they are holding the two ends of a pole without pulling or pushing:



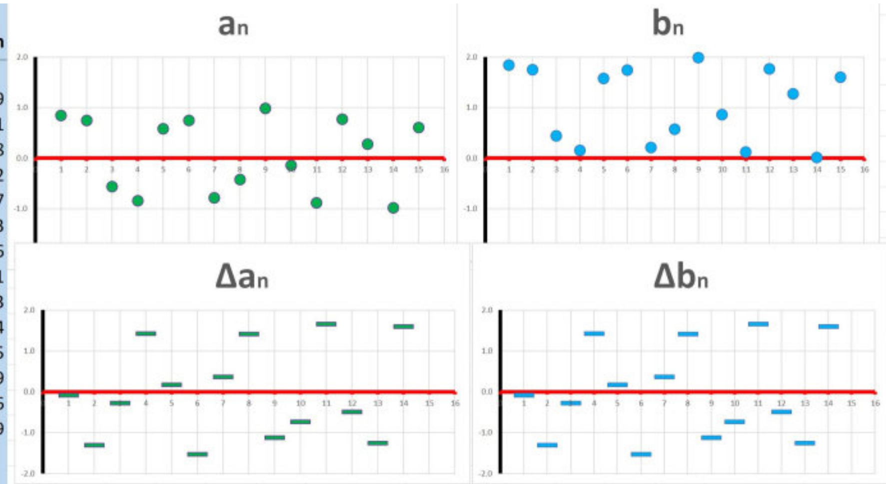
It is even possible that they speed up and slow down all the time. The velocity is the same because they move like a single body.

Once again, for functions $y=f(x)$ and $y=g(x)$ representing their position, we can restate this idea mathematically in order to confirm that our theory makes sense.

Example 3.7.8: shift of sequence

This is what the construction looks like for sequences. We shift the sequence a_n below by 1 unit up to produce a new sequence b_n (top):

n	a_n	b_n	Δa_n	Δb_n
1	0.84	1.84		
2	0.75	1.75	-0.09	-0.09
3	-0.56	0.44	-1.31	-1.31
4	-0.84	0.16	-0.28	-0.28
5	0.58	1.58	1.42	1.42
6	0.74	1.74	0.17	0.17
7	-0.79	0.21	-1.53	-1.53
8	-0.43	0.57	0.36	0.36
9	0.99	1.99	1.41	1.41
10	-0.14	0.86	-1.13	-1.13
11	-0.88	0.12	-0.74	-0.74
12	0.77	1.77	1.65	1.65
13	0.27	1.27	-0.49	-0.49
14	-0.99	0.01	-1.26	-1.26
15	0.60	1.60	1.59	1.59



Because the ups and downs remain the same, the sequences of differences of these two sequences are identical (bottom).

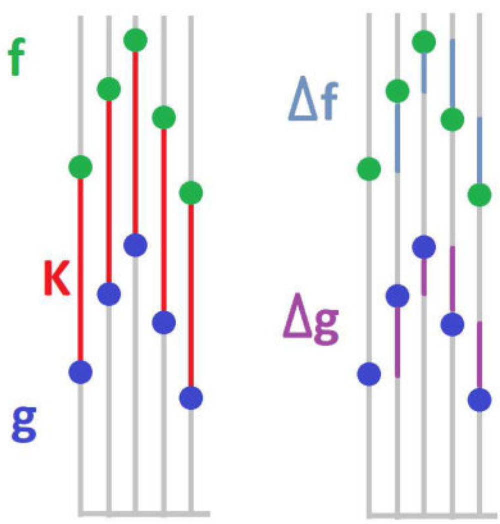
We follow what we know about the differences of sequences (seen in Volume 1, Chapter 1PC-2):

▶ If two functions defined at the nodes of a partition of interval $[a, b]$ differ by a constant, then they have the same differences.

In other words, we have the following:

$$f(x) - g(x) = K \implies \Delta f(c) = \Delta g(c)\,.$$

The functions below are colored accordingly:



In other words, if the height of a tunnel is constant, the floor and the ceiling must be identical in shape. Below is the algebraic representation of this idea:

Theorem 3.7.9: Difference of Functions That Differ by Constant

If two functions defined at the nodes of a partition of interval $[a, b]$ differ by a constant, then they have the same differences.

In other words, we have:

$$f(x) - g(x) = \text{ constant } \implies \Delta f(c) = \Delta g(c)\,.$$

We just divide by Δx to prove the following:

Theorem 3.7.10: Difference Quotient of Functions That Differ by Constant

If two functions defined at the nodes of a partition of interval $[a, b]$ differ by a constant, then they have the same difference quotient.

In other words, we have:

$$f(x) - g(x) = \text{ constant } \implies \frac{\Delta f}{\Delta x}(c) = \frac{\Delta g}{\Delta x}(c)\,.$$

Taking the limit $\Delta x \to 0$ proves the following:

Theorem 3.7.11: Derivative of Functions That Differ by Constant

If two differentiable on open interval I functions differ by a constant, then their derivatives are equal.

In other words, we have:

$$f(x) - g(x) = \text{ constant } \implies \frac{df}{dx}(x) = \frac{dg}{dx}(x)\,.$$

In summary:

$$f(x) - g(x) = \text{ constant } \implies \Delta f(c) = \Delta g(c) \implies \frac{\Delta f}{\Delta x}(c) = \frac{\Delta g}{\Delta x}(c) \implies \frac{df}{dx}(x) = \frac{dg}{dx}(x)\,.$$

Here, we move from displacements to average velocities to instantaneous velocities.

Exercise 3.7.12

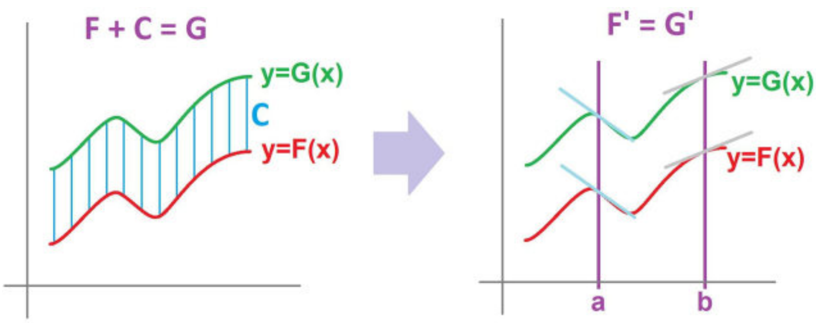
Provide details of the proof.

However, if we are willing to use *Derivative of Constant Function*, this proof might be much shorter:

$$f(x) - g(x) = \text{ constant } \implies \frac{d(f-g)}{dx}(x) = 0 \implies \frac{df}{dx}(x) = \frac{dg}{dx}(x)\,.$$

The last implication is to be proved shortly.

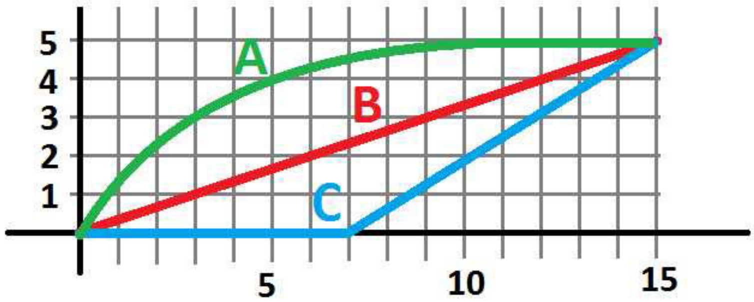
The conclusion can be confirmed by looking at the graphs:



If they differ by a constant, one is just a vertically shifted version of the other. Therefore, the graphs have exactly the same shape. Therefore, the graphs have exactly the same slopes at the corresponding locations. Therefore, the derivatives are exactly the same.

Example 3.7.13: three runners

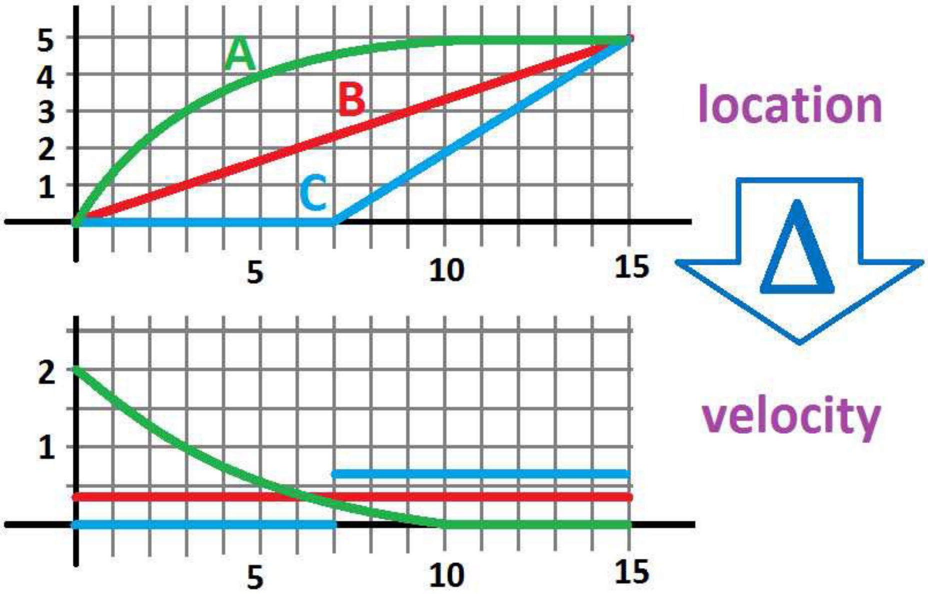
The graph shows the positions of three runners as functions of time. Describe what has happened.



Here's what happened:

- Runner A starts fast and then slows down, but reaches the finish line first.
- Runner B maintains the same speed.
- Runner C starts late and then runs fast and arrives at the same time as B.

We estimate the slopes of the graph and discover several values of the velocities. The three graphs are sketched here:

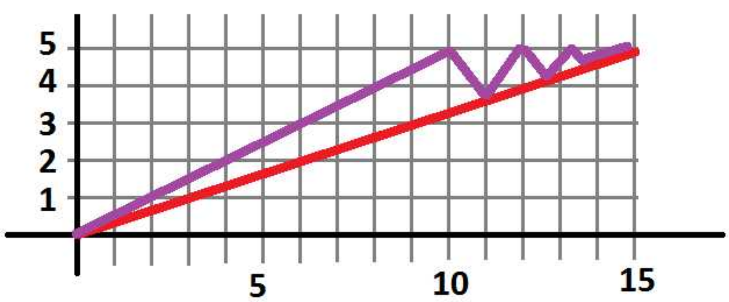


Exercise 3.7.14

Find the exact values of the constant velocities.

Exercise 3.7.15

Describe what has happened here:



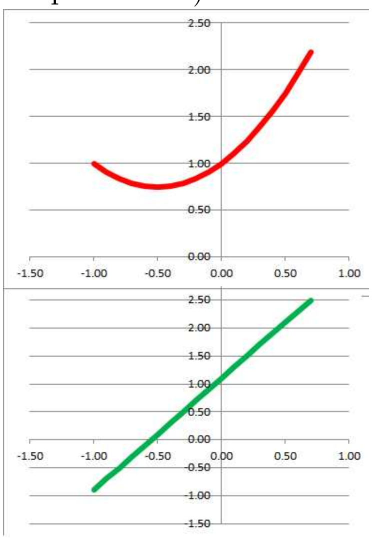
Example 3.7.16: general quadratic function

What if we pick a quadratic function this time? Let's find $f' = \frac{df}{dx}$ for $f(x) = ax^2 + bx + c$.

$$\begin{aligned}\frac{\Delta f}{\Delta x}(x) &= \frac{f(x+h) - f(x)}{h} \\ &= \frac{(a[x+h]^2 + b[x+h] + c) - (ax^2 + bx + c)}{h} \\ &= \frac{ax^2 + 2axh + ah^2 + bx + bh + c - ax^2 - bx - c}{h} && \text{The terms without } h \text{ cancel.} \\ &= \frac{2axh + ah^2 + bh}{h} && \text{That's why we can divide by } h! \\ &= 2ax + ah + b\,. \quad \Longrightarrow \\ \frac{df}{dx}(x) &= \lim_{h\to 0}(2ax + ah + b) \\ &= (2ax + ah + b)\Big|_{h=0} \\ &= 2ax + b\,.\end{aligned}$$

As we can see, the difference between the difference quotient and the derivative is a small vertical shift.

We can now match the corresponding parts of the graphs of the function and its derivative (the difference quotient is, however, what we plot below):



We see how the areas of the graph of f with positive/negative *slopes* correspond to the area of the graph of $f' = \frac{df}{dx}$ with positive/negative *values*.

The example proves the first half of the following statement:

▶ The derivative of a quadratic function is linear and, conversely, a linear function can only be the derivative of a quadratic function.

The second half will remain a conjecture for now.

Exercise 3.7.17

Prove that the derivative of an odd function is even, and the derivative of an even function is odd.

Example 3.7.18: rolling ball

Suppose we have a rolling ball that hits a wall and then goes back.

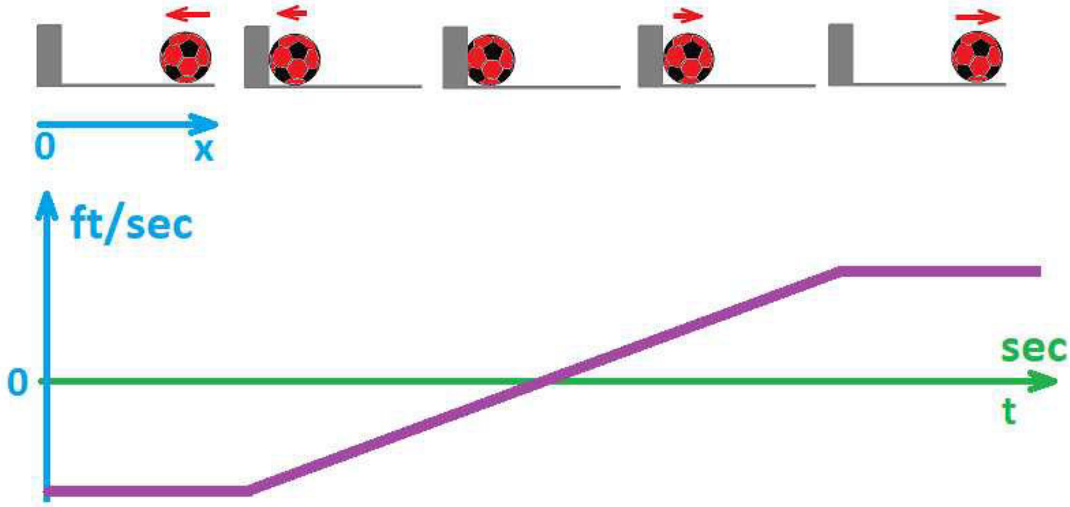
This time, the motion is described in terms of its velocity rather than its location! We want to find the location – of the center of the ball – as a function of time.

First, we add the x-axis to the picture. We make the positive direction to be to the right of the wall. Then we observe the following about the velocity:

1. The velocity of the ball is initially negative and constant.
2. As the ball touches the wall and starts to contract, the speed declines. This means that the velocity increases!
3. The velocity increases until it reaches 0.
4. The velocity then continues to increase as the ball starts to expand.
5. Finally, as the ball leaves the wall, the velocity that has been reached becomes constant and positive.

These are plausible values:

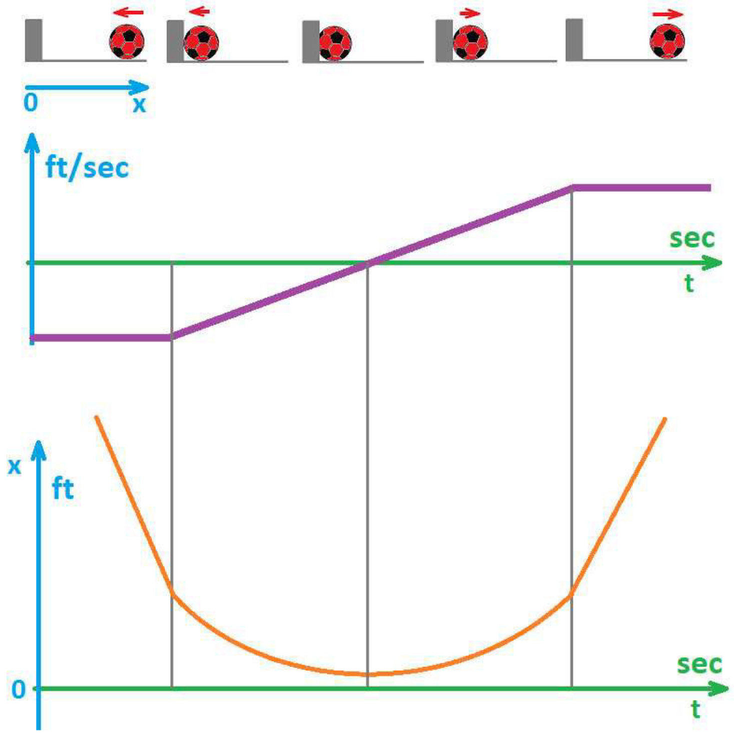
We assume, naturally, that the velocity is changing continuously. Then we connect these values into a curve following the description above:



Next, based on the graph of the velocity, we plot the location. Here we use the facts established above:

- The position is a linear location function when the velocity is constant.
- The position is a quadratic location function when the velocity is linear.

Furthermore, we assume, very naturally, that the location is changing continuously. We then connect the linear and quadratic pieces into a single curve:



Exercise 3.7.19

Does the graph of location touch the t-axis? Does the graph of location have cusps at the moments when the ball reaches and leaves the wall?

Exercise 3.7.20

What if this is a billiard ball and the collision is perfectly rigid? Plot the resulting functions and discuss their continuity and differentiability.

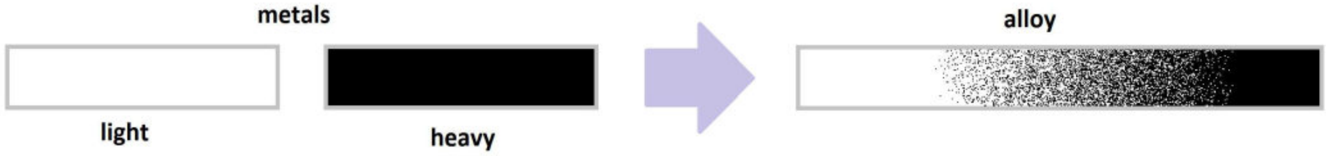
Example 3.7.21: linear density

Suppose we are given a uniform metal rod:

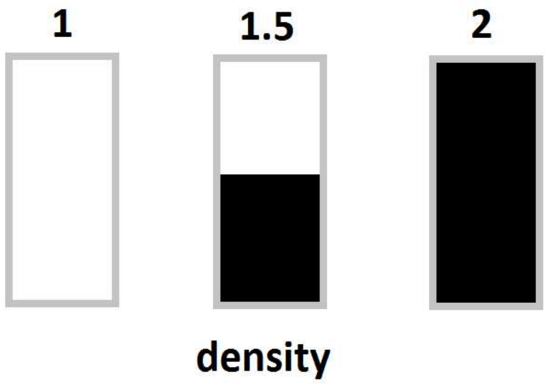
We define for such a rod:

$$\text{average linear density} = \frac{\text{mass}}{\text{length}}.$$

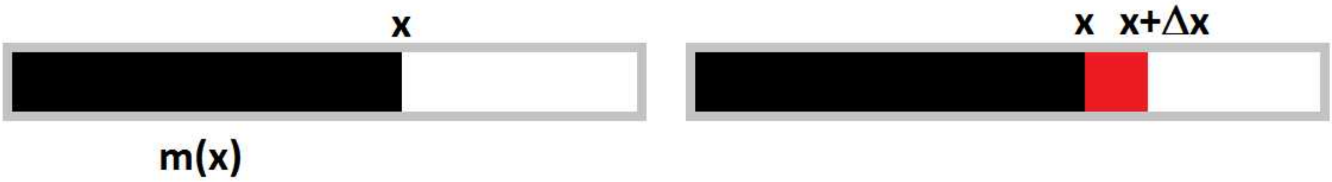
What if the rod isn't uniform, like an alloy? For example, we might have two pieces of different metals partially melted together:



Suppose the densities are 1 lb/in and 2 lbs/in at the ends respectively. Then the density of this alloy will gradually change from 1 to 2:



This is the *mass function* of x, the location: $y = m(x)$ is the mass of the rod from 0 to x. Consider how m is changing:



We are in the same position we were when we moved from constant to variable velocities!

Take a small piece of the rod at location x, Δx long, and let's call its mass Δm. Then, for this piece, we have:

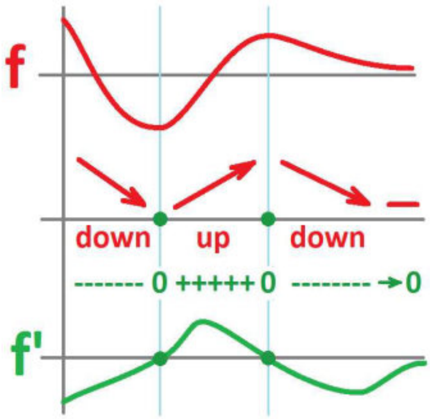
$$\text{average density} = \frac{\text{mass}}{\text{length}} = \frac{\Delta m}{\Delta x} = \frac{m(x+\Delta x) - m(x)}{\Delta x}\,.$$

Then we define the linear density as the derivative of the mass function:

$$\text{linear density at } x = \lim_{\Delta x \to 0} \frac{\Delta m}{\Delta x} = m'(x)\,.$$

Example 3.7.22: sketch the derivative

We start with the graph of f in red (first row) and look for its positive and negative slopes (second row):



Of course, we are looking at the increasing and decreasing behavior. We write $+$ for increasing and $-$ for decreasing (third row). This information is sufficient to plot a rough graph of f' (last row).

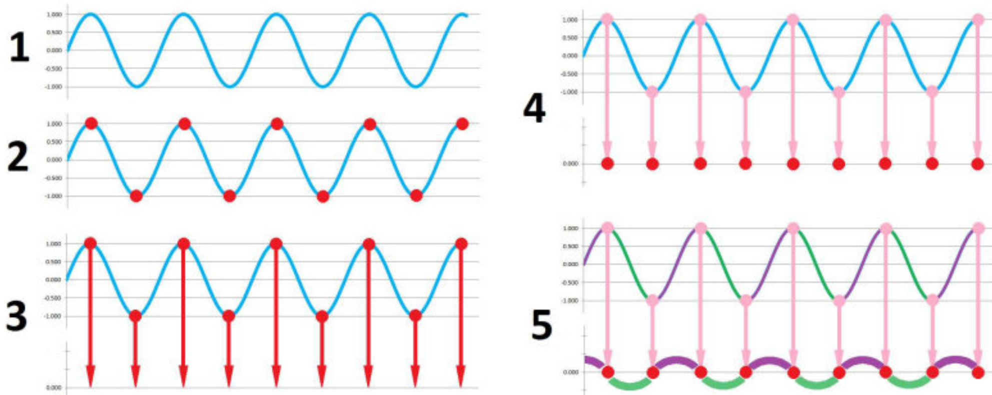
As a shortcut, we can first look at the points with a horizontal tangent, i.e., for those x's where $f'(x) = 0$. These will typically separate intervals where $f' > 0$ from $f' < 0$. We also notice the flattening of the graph on the far right; the result is also a flattening of the graph of the derivative – toward 0.

Warning!

Even though f' is *derived* from f, as you can see, there is no way to conjure the whole graph of f' from the looks of the graph of f. We shouldn't expect the graph of f' to emerge from shifting the graph of f, stretching or shrinking, or flipping, etc.

Example 3.7.23: qualitative analysis

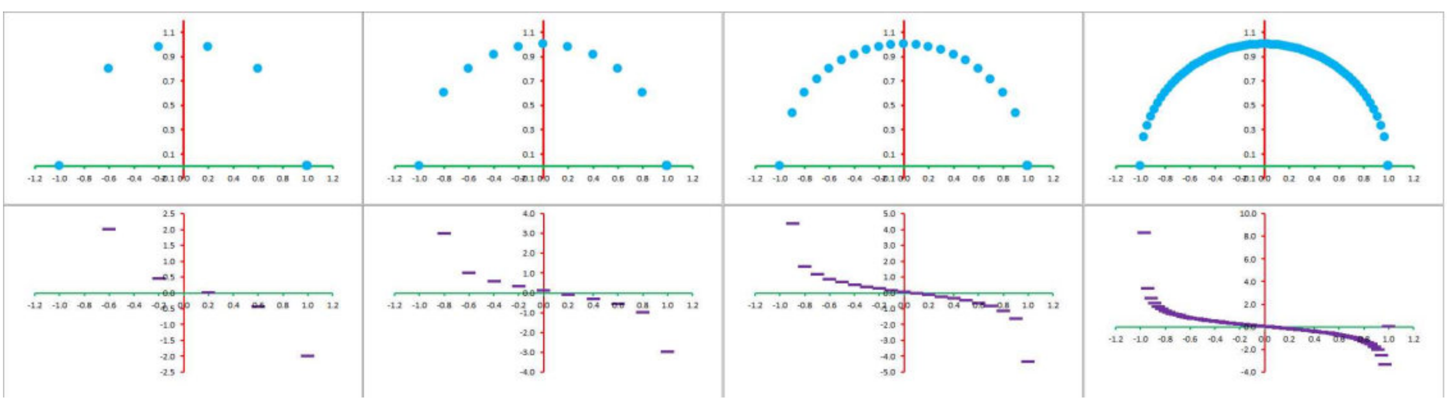
We can do this more strategically, with fewer points to test. Let's apply this method to the graph of $y = \sin x$. It is repetitive:



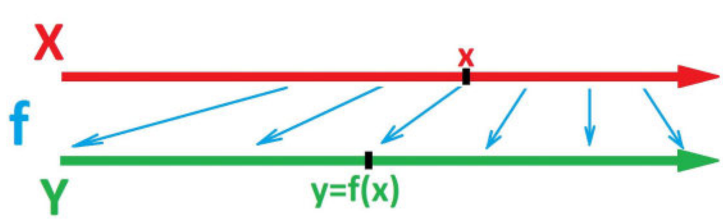
The resulting graph of the derivative looks like that of $y = \cos x$! We will show below that this isn't a coincidence.

Let's summarize.

As Δx is approaching zero, there are more and more nodes in the partition and more and more points found on the graph of the function. As a result, there are more and more values of the difference quotient found. Eventually, the former points form the graph of the function and the latter form the graph of its derivative:

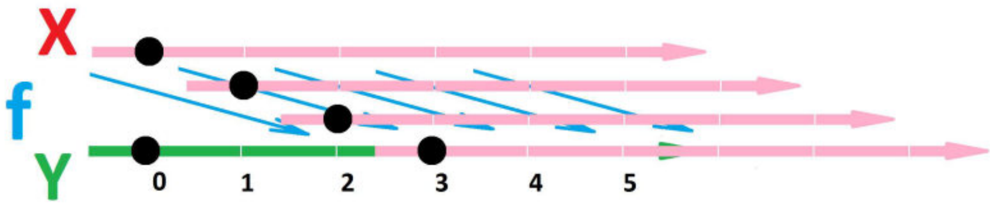


Next, we consider the meaning of the derivative of a function that represents a *transformation*:



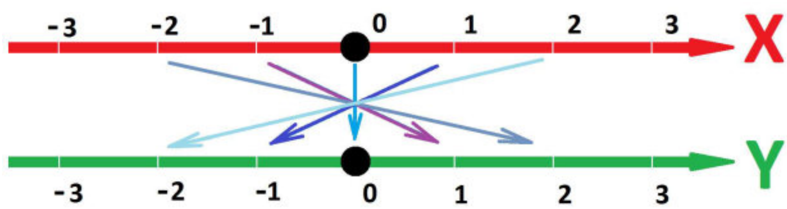
Let's start with the basic transformations.

First a shift $y = x + k$:



The derivative is 1.

Next, a flip $y = -x$:

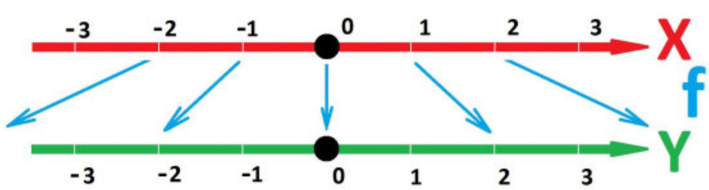


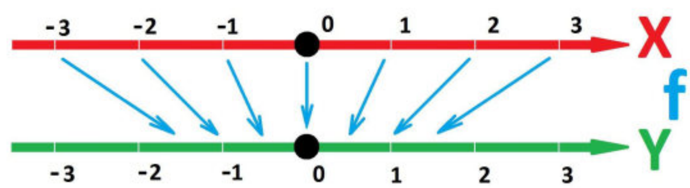
The derivative is -1.

We conclude the following:

- ▶ For rigid motion (no changes of distances), the derivative is 1 or -1.

Next, a stretch $y = kx$ ($k > 0$):

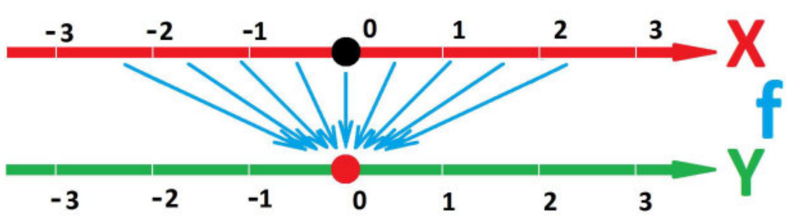




The derivative is k. The stretch $y = 2x$ is uniform as the distance between *any* two points doubles. It is also the derivative! In general, this ratio k is the ratio of the lengths of the corresponding segments in the range and the domain.

Furthermore, a shrink is a stretch with $k < 1$.

The collapse, $f(x) = c$, makes all distances equal to 0:

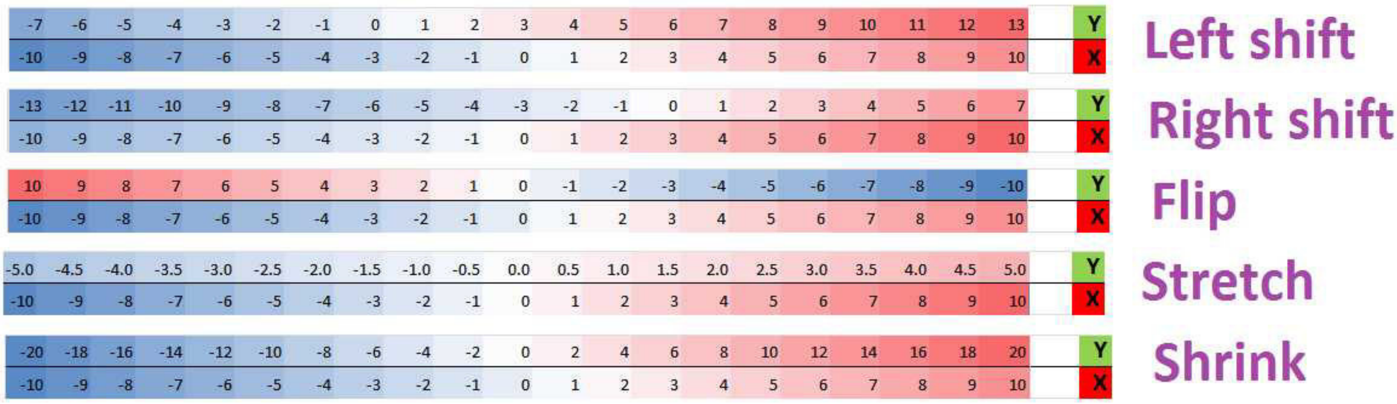


Therefore, its derivative is 0 too.

We conclude:

- ▶ The derivative describes how the *relative locations* of points change under the transformation.

This is the summary of the functions we have considered:



Left shift

-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12	13	Y
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	X

Right shift

-13	-12	-11	-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	Y
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	X

Flip

10	9	8	7	6	5	4	3	2	1	0	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	Y
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	X

Stretch

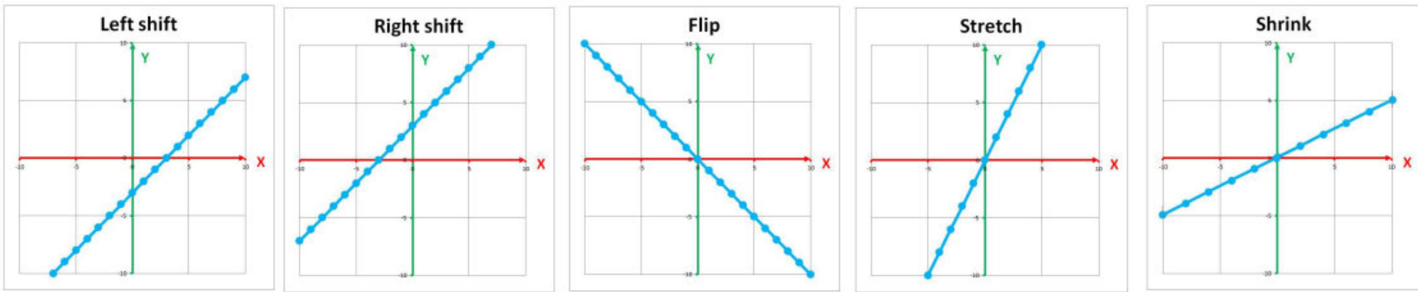
-5.0	-4.5	-4.0	-3.5	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0	Y
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	X

Shrink

-20	-18	-16	-14	-12	-10	-8	-6	-4	-2	0	2	4	6	8	10	12	14	16	18	20	Y
-10	-9	-8	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	X

Their derivatives are respectively: 1, 1, -1, 2, and $1/2$. They are also the stretching factors!

We also plot the graphs of these functions below to see these factors as the slopes:



We know (from Volume 1, Chapter 1PC-3) that under a linear polynomial $f(x) = mx + b$, the distances increase by a factor of $|m|$ when $|m| > 1$, or decreased by a factor of $|m|$ when $|m| < 1$. This stretch/shrink factor is the same everywhere. But m is the derivative of f!

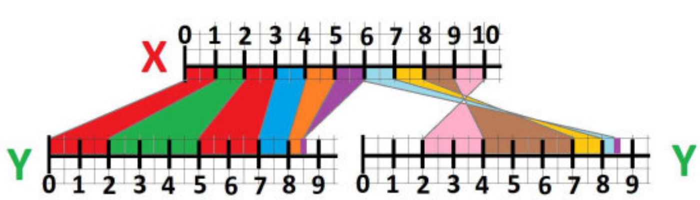
Example 3.7.24: transformations

Let's consider this function given by its values:

x	0	1	2	3	4	5	6	7	8	9	10
y	0	2	5	7	8	8.5	8.5	8	7	4	2

We assume that the function continues between these values in a linear fashion. For example, the interval $[0, 1]$ is mapped to $[0, 2]$ linearly ($y = 2x$), the interval $[1, 2]$ to $[2, 5]$, etc. The actual formulas don't matter.

Then, the 1-unit segments on the x-axis are stretched and shrunk at different rates, and the ones beyond $x = 6$ are also flipped over (left):



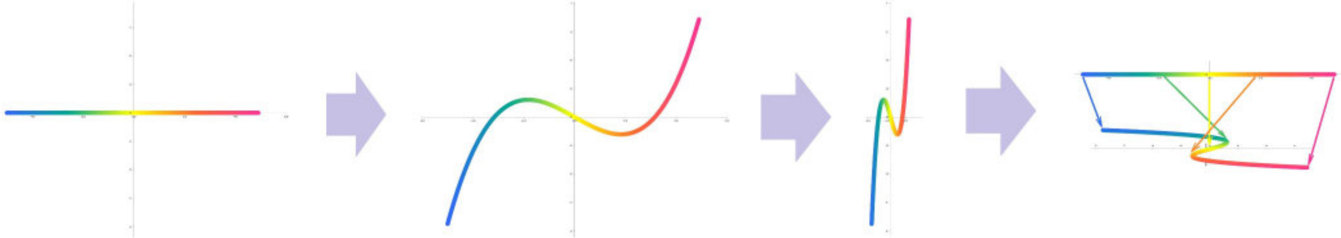
The derivatives are respectively:

$$2, \ 3, \ 2, \ 1, \ 1/2, \ 0, \ -1/2, \ -1, \ -3 \,.$$

So, the absolute values of these numbers are the stretch coefficients of these segments of the domain. We use these factors as slopes for the patches that make up the graph (right).

Exercise 3.7.25

Estimate the derivative from the fourth picture and then confirm the result by looking at the slopes in the second:



In summary, an abstract numerical function $y = f(x)$ has been given three *tangible representations* and now we also have three tangible interpretations of its derivative:

	Function is seen as	Its derivative is
1.	location	velocity
2.	graph	slope of the tangent line
3.	transformation	stretch/shrink rate

3.8. Basic differentiation

Let's review our current, and then introduce a new, notation.

Initially, we deal with one point at a time, $x = a$. First, the difference quotient:

$$\frac{\Delta f}{\Delta x}(a) = \frac{f(a + \Delta x) - f(a)}{\Delta x} = \frac{f(x) - f(a)}{x - a}$$

And this is the derivative:

$$\frac{df}{dx}(a) = \lim_{\Delta x \to 0} \frac{\Delta f}{\Delta x}(a)$$

We can omit the mention of the input variable and deal with only the names of these functions:

$$\frac{\Delta f}{\Delta x} \text{ and } \frac{df}{dx}$$

Now, it is also possible to leave the original function unnamed, as in:

$$y = x^2 .$$

In that case, we may use the name of the output variable of this function for the difference quotient and the derivative:

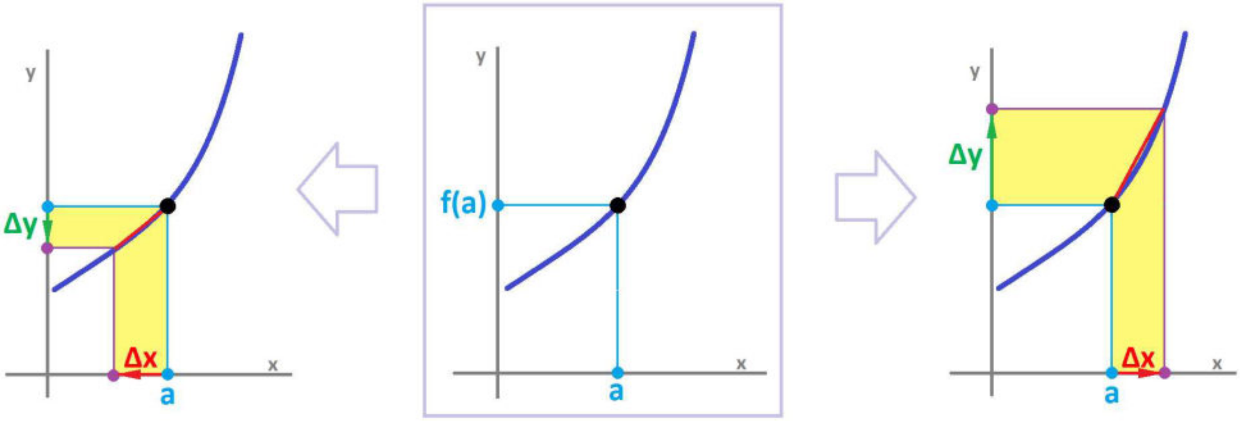
Difference quotient and derivative

$$\frac{\Delta y}{\Delta x} \text{ and } \frac{dy}{dx}$$

In the difference quotient, we use the Greek letter Δ, which stands for "difference", as follows:

- Δx is the change of x (the run), and
- Δy is the change of y (the rise).

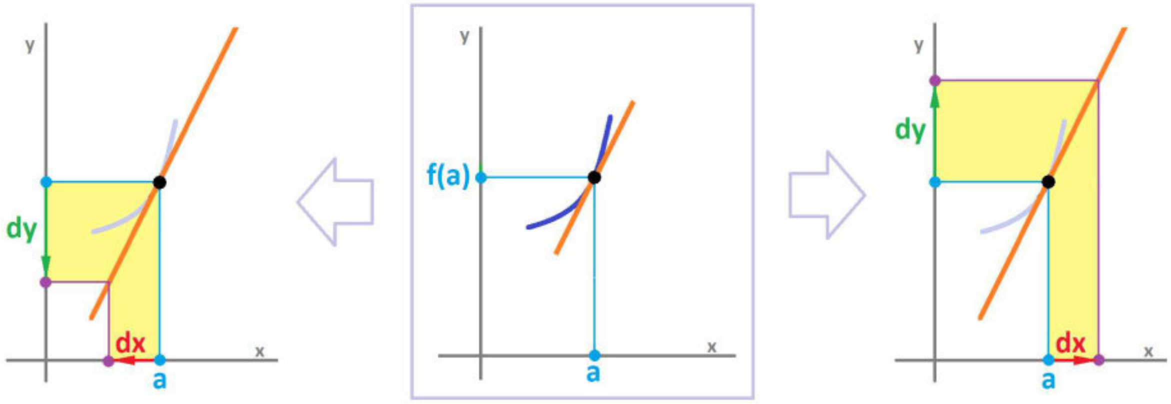
The former produces the latter with the use of the graph of $y = f(x)$. For every choice of Δx, we make that step to the left or to the right and find another point on the graph:



Next, in the derivative, we use the letter d that, too, stands for "difference". The derivative is the limit of a fraction, but this doesn't mean that – in spite of the notation – it is a fraction too! However, once the derivative is known, at $x = a$, it gives us the slope of the tangent line. Just as above, we concentrate on a single line – the tangent line – and use this notation:

- dx for the run, and
- dy for the rise.

For every choice of dx, we make that step to the left or to the right and find another point on the tangent line:



So, the derivative is a fraction, too:

$$\frac{dy}{dx} = \frac{\text{change of } y}{\text{change of } x} .$$

If in the above picture, the slope is 2, we have a relation between these two variables:

$$dy = 2\, dx .$$

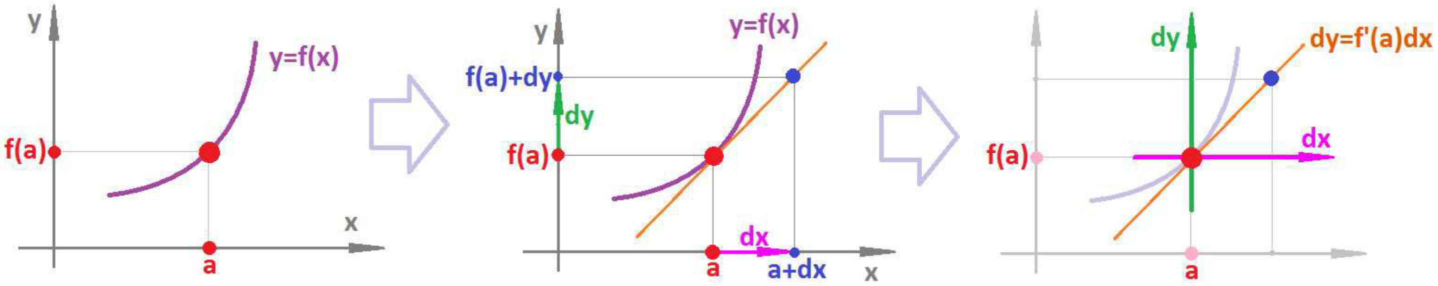
We treat these two quantities, dx and dy, as a *new set of variables.* They depend on each other, as follows:

Definition 3.8.1: differentials

Suppose a function $y = f(x)$ is differentiable at $x = a$ and its derivative is $f'(a)$. Then the *differential*, dx, of x and the *differential*, dy, of y are two real variables related to each other by the equation:

$$\boxed{dy = f'(a) \cdot dx}$$

Such an expression is called a *differential form* (to be considered in Volume 3, Chapters 3IC-2 and 3IC-4).



At every location a, the dependence of the two variables is very simple: linear!

Another approach to notation is to present differentiation as a function, *a function of functions*:

Difference quotient and derivative

$$\frac{\Delta}{\Delta x}(f) \quad \text{and} \quad \frac{d}{dx}(f)$$

Here both the input and the output are functions!

Also convenient sometimes is the *substitution notation* for evaluating a function at a particular point:

$$\frac{\Delta f}{\Delta x}\bigg|_{x=a}, \ \frac{\Delta y}{\Delta x}\bigg|_{x=a}, \ \frac{df}{dx}\bigg|_{x=a}, \ \frac{dy}{dx}\bigg|_{x=a}.$$

Example 3.8.2: computations

We may, in fact, omit the names altogether and present only the formula of the function according to the theorems in the last section. This is what we write for the difference quotients:

$$\begin{array}{llll}
\frac{\Delta}{\Delta x}(3x+5)=3 & \Longrightarrow \frac{\Delta}{\Delta x}(3x+5)\bigg|_{x=0} & =3\bigg|_{x=0} & =3\\
\frac{\Delta}{\Delta x}(-x^2+7)=-2x-2h & \Longrightarrow \frac{\Delta}{\Delta x}(-x^2+7)\bigg|_{x=1,\ h=.1} & =-2x-2h\bigg|_{x=1,\ h=.1} & =-2.2
\end{array}$$

This is what we write for the derivatives:

$$\begin{array}{llll}
\frac{d}{dx}(-x^2+7)=-2x & \Longrightarrow \frac{d}{dx}(-x^2+7)\bigg|_{x=1} & =-2x\bigg|_{x=1} & =-2\\
\frac{d}{dx}(3x^2+2x+1)=6x+2 & \Longrightarrow \frac{d}{dx}(3x^2+2x+1)\bigg|_{x=-1} & =6x+2\bigg|_{x=-1} & =-4
\end{array}$$

And now in the Lagrange notation:

$$\begin{array}{llll}
(-x^2+7)'=-2x & \Longrightarrow (-x^2+7)'\bigg|_{x=1} & =-2x\bigg|_{x=1} & =-2\\
(3x^2+2x+1)'=6x+2 & \Longrightarrow (3x^2+2x+1)'\bigg|_{x=-1} & =6x+2\bigg|_{x=-1} & =-4
\end{array}$$

Warning!

We can't substitute $x = 1$ into $(-x^2 + 7)'$ until an explicit representation of the latter is found.

Exercise 3.8.3

Compute:

$$\begin{array}{ll} \frac{\Delta}{\Delta x}(-x+2) & \text{at } x = 5 \\ \frac{d}{dx}(-x+2) & \text{at } x = 5 \\ \frac{d}{dx}(e^x) & \text{at } x = -1 \\ (\sin x)' & \text{at } x = \pi/2 \end{array}$$

Exercise 3.8.4

Prove plotting the difference quotient of a quadratic function will produce a straight line.

Example 3.8.5: spreadsheet

Computing might give us evidence to make an educated guess about the derivatives of particular functions.

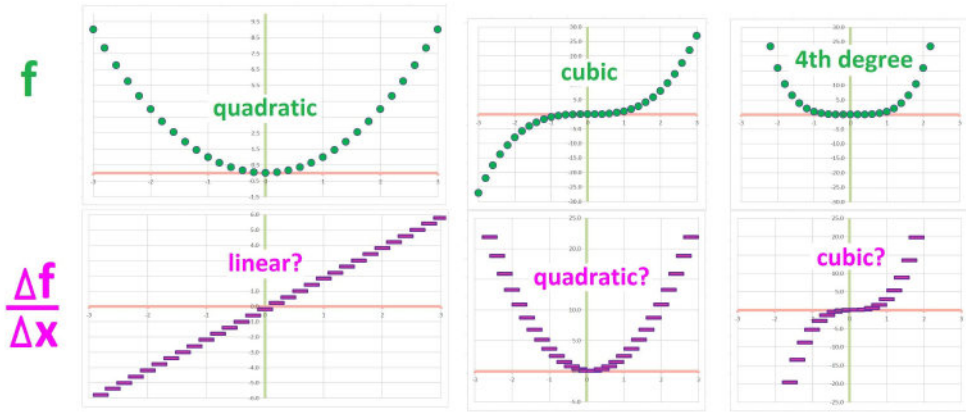
Consider a parabola, say, $y = f(x) = -(x - 1.5)^2 + 3$. Recall how in search of a pattern, we make the nodes of our partition denser and denser:

h=	1.00		
n	x	y	DQ
1	-3.0	-17.250	
2	-2.0	-9.250	8.0
3	-1.0	-3.250	6.0
4	0.0	0.750	4.0
5	1.0	2.750	2.0
6	2.0	2.750	0.0
7	3.0	0.750	-2.0

h=	0.05		
n	x	y	DQ
1	-3.0	-17.250	
2	-3.0	-16.803	8.9
3	-2.9	-16.360	8.8
4	-2.9	-15.923	8.8
5	-2.8	-15.490	8.6
6	-2.8	-15.063	8.5
7	-2.7	-14.640	8.4
8	-2.7	-14.223	8.3
9	-2.6	-13.810	8.2
10	-2.6	-13.403	8.1
11	-2.5	-13.000	8.1
12	-2.5	-12.603	7.9
13	-2.4	-12.210	7.8
14	-2.4	-11.823	7.7
15	-2.3	-11.440	7.6
16	-2.3	-11.063	7.5
17	-2.2	-10.690	7.4
18	-2.2	-10.323	7.3
19	-2.1	-9.960	7.3

As we produce more and more points on this interval, we realize that the difference quotient is a straight line!

We will see more patterns if we collect the difference quotients of the functions we have seen previously. These are those of the power functions:



It appears that the power goes down by one!

Exercise 3.8.6

Sketch the next pair.

Example 3.8.7: x^2

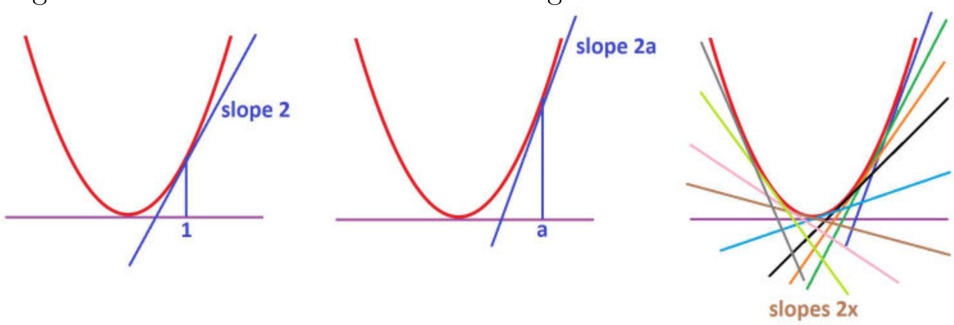
We have made progress with the derivative of

$$f(x) = x^2 .$$

From a single specific point, to a single unspecified point, to all points at once, a new function! However, the algebraic steps are the same:

$a = 1$	replace 1 with a	replace a with x	
$\dfrac{\Delta f}{\Delta x}(1) =$	$\dfrac{\Delta f}{\Delta x}(a) =$	$\dfrac{\Delta f}{\Delta x}(x) =$	The difference quotient...
$= \dfrac{f(1+h)-f(1)}{h}$	$= \dfrac{f(a+h)-f(a)}{h}$	$= \dfrac{f(x+h)-f(x)}{h}$	is written from the definition.
$= \dfrac{(1+h)^2-1^2}{h}$	$= \dfrac{(a+h)^2-a^2}{h}$	$= \dfrac{(x+h)^2-x^2}{h}$	The function is specified.
$= \dfrac{1^2+2h+h-1}{h}$	$= \dfrac{a^2+2ah+h-a^2}{h}$	$= \dfrac{x^2+2xh+h-x^2}{h}$	The numerator is expanded.
$= \dfrac{2h+h^2}{h}$	$= \dfrac{2ah+h^2}{h}$	$= \dfrac{2xh+h^2}{h}$	The terms without h are canceled.
$= 2+h$	$= 2a+h$	$= 2x+h$	The numerator is divided by h.
$\to 2+0$	$\to 2a+0$	$\to 2x+0$	The limit is evaluated by...
$= 2$	$= 2a$	$= 2x$	substitution because the function...
			is continuous with respect to h.

Our progress is illustrated in terms of the tangent lines:



In this section, we will find the derivatives of some important functions. The functions are very different, but the computations will have a lot in common.

We need to find the limit of the difference quotient,

$$\frac{\Delta f}{\Delta x} = \frac{f(x+h) - f(x)}{h} \quad \text{as} \quad h \to 0\,.$$

Whenever f is continuous at x, the limit of the numerator is 0. And so is the limit of the denominator! Thus, we will face, every time, the same problem: an indeterminate expression of the type $\frac{0}{0}$. Every time, it is to be *resolved*, and not by the rules of limits but by algebra! This is the most challenging step:

- ▶ We will need to *factor* the numerator in order to cancel h from the difference quotient.

Example 3.8.8: expanding difference quotient of x^3

The *power functions* first. We already know the derivative for the linear and the quadratic powers:

$$(x^1)' = 1,\ (x^2)' = 2x^1\,.$$

Now cubic:

$$f(x) = x^3\,.$$

We follow the approach in the last section, keeping in mind that x is fixed as far as the limit is concerned:

$\frac{\Delta f}{\Delta x}(x)$	$= \frac{f(x+h) - f(x)}{h}$	The difference quotient is written from the definition.
	$= \frac{(x+h)^3 - x^3}{h}$	The function is specified.
	$= \frac{x^3 + 3x^2h + 3xh^2 + h^3 - x^2}{h}$	The numerator is expanded.
	$= \frac{3x^2h + 3xh^2 + h^3}{h}$	The terms without h are cancelled.
	$= 3x^2 + 3xh + h^2$	The numerator is divided by h.
$\frac{\Delta f}{\Delta x}(x)$	$= 3x^2 + 3xh + h^2$	This is the simplified difference quotient.
as $h \to 0$,	$\to 3x^2 + 3x \cdot 0 + 0^2$	The limit is then evaluated by substitution $h = 0$...
	$= 3x^2$	because the expression is continuous with respect to h.

We notice something: All the terms disappear except for those with h in the *first power*.

The following is derived from the *Binomial Formula* (seen in Volume 1, Chapter 1PC-1):

Theorem 3.8.9: Highest Terms of Binomial Expansion

The nth power, $(x+h)^n$, has $n+1$ terms, and the one with h has coefficient equal to n:

$$(x+h)^n = x^n + nx^{n-1}h + \text{ terms with } h^2,\ h^3,\ \ldots$$

We will use this fact later.

An alternative approach to factoring relies on the following *factoring formula*:

$$\boxed{a^n - b^n = (a-b)(a^{n-1} + a^{n-2}b + \ldots + ab^{n-2} + b^{n-1})}$$

Example 3.8.10: factoring difference quotient of x^3

We use it as follows ($n=3$):

$$\begin{aligned}(x^3)' &= \lim_{h\to 0}\frac{(x+h)^3-x^3}{h}\\ &= \lim_{h\to 0}\frac{(x+h-x)((x+h)^2+(x+h)x+x^2)}{h}\\ &= \lim_{h\to 0}\left[(x+h)^2+(x+h)x+x^2\right]\\ &= (x+0)^2+(x+0)x+x^2\\ &= 3x^2\,.\end{aligned}$$

We have also discovered a formula for the difference quotient:

$$\frac{\Delta}{\Delta x}\left(x^3\right) = (x+h)^2+(x+h)x+x^2\,.$$

We plot all three below:

A pattern starts to emerge:

n	$(x^n)'$				
1	$(x^1)'$	$=1x^0$	1	x	0
2	$(x^2)'$	$=2x^1$	2	x	1
3	$(x^3)'$	$=3x^2$	3	x	2
...	...	...	...	...	
1000	$(x^{1000})'$	$\stackrel{?}{=} 1000x^{999}$	1000	x	999
...	...	...	...	...	
n	$(x^n)'$	$\stackrel{?}{=} nx^{n-1}$	n	x	$n-1$

It seems that the power present in the derivative is one lower than that in the original, while the original power appears as a multiple.

Let's go over the computation, starting with the difference:

Theorem 3.8.11: Integer Power Formula for Difference

Let n be a positive integer. Suppose we have a left-end partition; i.e., the nodes are $x = a,\ a + h$ and the secondary node is $c = a$. Then the difference of $y = x^n$ is given at c by the formula:

$$\Delta(x^n) = h\Big((c+h)^{n-1} + (c+h)^{n-2}c^1 + ... + (c+h)^1 c^{n-2} + c^{n-1}\Big).$$

Proof.

The proof only needs one of the two formulas above.

Now we just divide by $h = \Delta x$:

Theorem 3.8.12: Integer Power Formula for Difference Quotient

Let n be a positive integer. Suppose we have a left-end partition; i.e., the nodes are $x = a,\ a + h$ and the secondary node is $c = a$. Then the difference quotient of $y = x^n$ is given at c by the formula:

$$\frac{\Delta}{\Delta x}(x^n) = (c+h)^{n-1} + (c+h)^{n-2}c^1 + ... + (c+h)^1 c^{n-2} + c^{n-1}.$$

Now we just let $h = \Delta x \to 0$. Then each term above becomes c^{n-1}. There are n of them:

Theorem 3.8.13: Integer Power Formula for Derivative

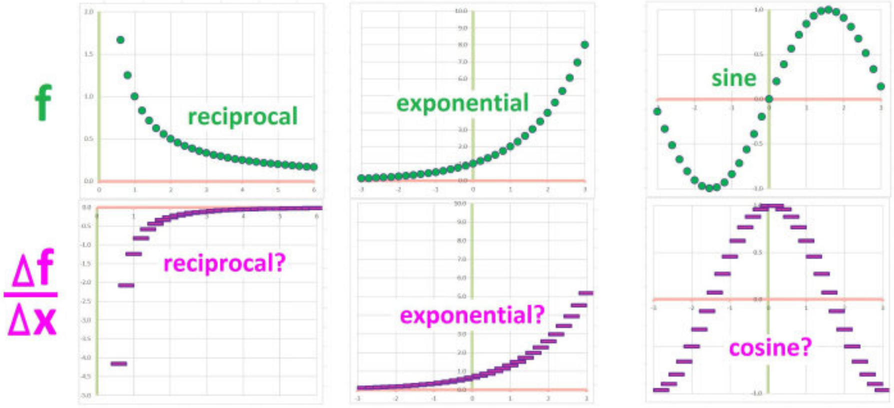
Let n be a positive integer. The derivative of $y = x^n$ is given by

$$\boxed{\frac{d}{dx}(x^n) = nx^{n-1}}$$

The limit dramatically simplifies the formula!

3.9. Basic differentiation, continued

There are more apparent patterns to be confirmed:



These are our simplest guesses. For the first one, the graph of the difference quotient seems to be the

negative and upside-down version of that of f; that's why our first guess is $-1/x$. However, the second graph is approaching the x-axis faster! Maybe it's $-1/x^2$? Yes.

Exercise 3.9.1

Try to answer these questions.

The *Integer Power Formula* is now to be tested for other values of n.

Example 3.9.2: reciprocal

Let's now try negative powers. Compute:

$$\begin{aligned}
\left(\frac{1}{x}\right)' &= \lim_{h\to 0}\frac{\frac{1}{x+h}-\frac{1}{x}}{h} && \text{We do common denominators here.}\\
&= \lim_{h\to 0}\frac{\frac{x-(x+h)}{(x+h)x}}{h}\\
&= \lim_{h\to 0}\frac{-h}{(x+h)x}\frac{1}{h} && \text{Cancel } h.\\
&= \lim_{h\to 0}\frac{-1}{(x+h)x} && \text{The denominator is continuous for } h<|x|.\\
&= -\frac{1}{x^2} && \text{That's why we can just substitute } h=0.
\end{aligned}$$

How does this fit into our formula? We rewrite:

$$\frac{1}{x}=x^{-1},\ \frac{1}{x^2}=x^{-2},\ldots$$

So,

$$(x^{-1})'=-x^{-2}\,.$$

In other words, we have $n=-1$, and the formula still works. We have also discovered:

$$\frac{\Delta}{\Delta x}\left(x^{-1}\right)=\frac{-1}{(x+h)x}\,.$$

We plot all three below:

0.01	
x	f(x)
0.20	5.00
0.21	4.81
0.22	4.63
0.22	4.46
0.23	4.31
0.24	4.17
0.25	4.03
0.26	3.91
0.26	3.79

h=	0.20		
	x	f(x)	Δf/Δx
	0.20	5.00	
	0.40	2.50	-12.50
	0.60	1.67	-4.17
	0.80	1.25	-2.08
	1.00	1.00	-1.25
	1.20	0.83	-0.83
	1.40	0.71	-0.60
	1.60	0.63	-0.45
	1.80	0.56	-0.35
	2.00	0.50	
	2.20	0.45	
	2.40	0.42	
	2.60	0.38	
	2.80	0.36	
	3.00	0.33	
	3.20	0.31	
	3.40	0.29	
	3.60	0.28	
	3.80	0.26	
	4.00	0.25	
	4.20	0.24	
	4.40	0.23	
	4.60	0.22	
	4.80	0.21	

x	df/dx
0.20	-25.00
0.21	-23.11
0.22	-21.43
0.22	-19.93
0.23	-18.58
0.24	-17.36
0.25	-16.26
0.26	-15.26
0.26	-14.35

Example 3.9.3: square root

Next, fractional powers. Compute:

$$
\begin{aligned}
\left(\sqrt{x}\right)' &= \lim_{h\to 0} \frac{\sqrt{x+h}-\sqrt{x}}{h} && \text{Is this indeterminate?}\\
&= \lim_{h\to 0} \frac{\sqrt{x+h}-\sqrt{x}}{h}\cdot\frac{\sqrt{x+h}+\sqrt{x}}{\sqrt{x+h}+\sqrt{x}} && \text{This is the “rationalization trick”!}\\
&= \lim_{h\to 0} \frac{(x+h)-x}{h(\sqrt{x+h}+\sqrt{x})}\\
&= \lim_{h\to 0} \frac{1}{\sqrt{x+h}+\sqrt{x}} && \text{Because the function is continuous for } h<|x|.\\
&= \left.\frac{1}{\sqrt{x+h}+\sqrt{x}}\right|_{h=0} && \text{We substitute } h=0\,.\\
&= \frac{1}{2\sqrt{x}}\,.
\end{aligned}
$$

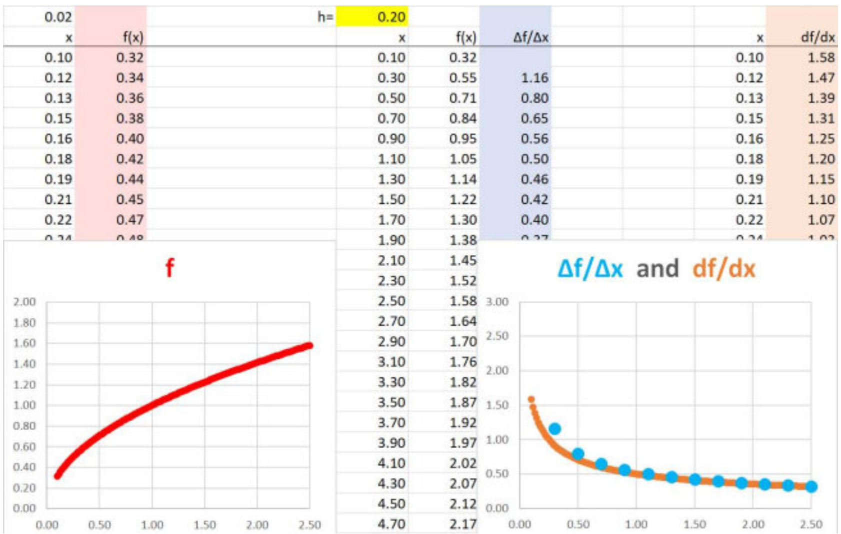
How does this fit into our formula? We rewrite:

$$\sqrt{x}=x^{1/2},\quad \frac{1}{\sqrt{x}}=x^{-1/2},\ldots$$

In other words, we have $n=1/2$, and the formula remains valid. We have also discovered:

$$\frac{\Delta}{\Delta x}\left(x^{1/2}\right)=\frac{1}{\sqrt{x+h}+\sqrt{x}}\,.$$

We plot all three below:



We also notice that:

$$f'(x)\to+\infty \quad\text{as}\quad x\to 0^+\,.$$

It follows that the graph of f becomes vertical at $x=0$. The function is not differentiable here.

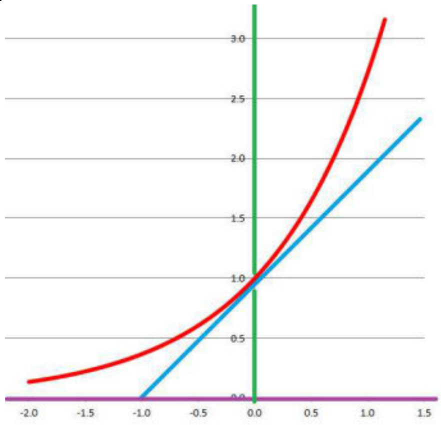
The general case, to be proven later, is as follows. For any real number $r\neq 0$, we have:

$$\boxed{(x^r)'=rx^{r-1}}$$

We will next compute the derivative of the *exponential function*.

Example 3.9.4: derivative of e^x at 0

Before addressing the general case, let's consider its derivative at $x = 0$. We know that the graph of $y = e^x$ crosses the y-axis at 45 degrees:



This means that the following famous limit:

$$\lim_{x\to 0} \frac{e^x - 1}{x} = 1\,,$$

has now a new interpretation:

$$(e^x)'\Big|_{x=0} = 1\,.$$

We now use this result to find the derivative function of the *exponential function* base e. Remarkably, the derivative of the exponential function is itself:

Theorem 3.9.5: Difference of Exponential Function

Suppose we have a left-end partition; i.e., the nodes are $x = a, a + h$ and the secondary node is $c = a$. Then the difference of e^x is given at c by:

$$\Delta(e^x) = (e^h - 1) \cdot e^c\,.$$

Proof.

Let $f(x) = e^x$. We compute at c:

$$\begin{aligned}
\Delta f &= e^{a+h} - e^a \\
&= e^a e^h - e^a && \text{According to a formula. Then we factor.} \\
&= e^a (e^h - 1)\,. && \text{Here } a \text{ is our secondary node!}
\end{aligned}$$

Theorem 3.9.6: Difference Quotient of Exponential Function

Suppose we have a left-end partition; i.e., the nodes are $x = a, a + h$ and the secondary node is $c = a$. Then the difference quotient of e^x is given at c by:

$$\frac{\Delta}{\Delta x}(e^x) = \frac{e^h - 1}{h} \cdot e^c\,.$$

Proof.

Let $f(x) = e^x$. We compute at c:

$$\begin{aligned} \frac{\Delta f}{\Delta x} &= \frac{e^a(e^h - 1)}{h} \\ &= e^a \cdot \frac{e^h - 1}{h}\,. \end{aligned}$$

Theorem 3.9.7: Derivative of Exponential Function

The derivative of e^x is given by:

$$\boxed{\frac{d}{dx}(e^x) = e^x}$$

Proof.

Let $f(x) = e^x$. We have at c:

$$\begin{aligned} \frac{\Delta f}{\Delta x} &= e^a \cdot \frac{e^h - 1}{h} \\ &\to e^a \cdot 1 \quad \text{ as } h \to 0 \\ &= e^a\,. \end{aligned}$$

The last step is justified by the *Constant Multiple Rule* and the famous limit above.

Let's compare the formulas for the difference quotient of e^x and its derivative:

$$\begin{array}{lll} \text{(A)} & \dfrac{\Delta}{\Delta x}(e^x) & = \dfrac{e^h - 1}{h} \cdot \quad e^x \\ \text{(B)} & \dfrac{d}{dx}(e^x) & = \qquad\qquad e^x \end{array}$$

We know that

$$k = \frac{e^h - 1}{h} > 1\,.$$

So, the graph of the difference quotient differs from that of the derivative only by a vertical stretch by a factor k! As h is approaching 0, the stretch diminishes and the graph of the former is approaching the graph of the latter:

The formula in the Lagrange notation is as follows:

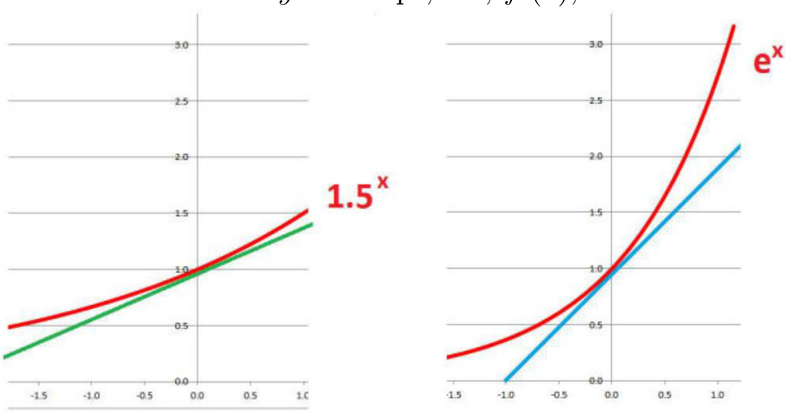
$$\boxed{(e^x)' = e^x}$$

Example 3.9.8: b^x

The above computation can be easily applied to the general exponential function, base b. Consider, $f(x) = b^x,\ b > 0$:

$$\begin{aligned}
(b^x)' = f'(x) &= \lim_{h\to 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h\to 0} \frac{b^{x+h} - b^x}{h} && \to \frac{0}{0}?\ \text{We need algebra!} \\
&= \lim_{h\to 0} \frac{b^x b^h - b^x}{h} && \text{We use a rule of exponents.} \\
&= \lim_{h\to 0} \frac{b^x(b^h - 1)}{h} && \text{Then factor and apply CMR.} \\
&= b^x \lim_{h\to 0} \frac{b^h - 1}{h} && \text{Does this limit exist?} \\
&= b^x \lim_{h\to 0} \frac{b^{0+h} - b^0}{h} && \text{It is a familiar one!} \\
&= b^x \cdot (b^x)' \Big|_{x=0}. && \text{No simplification here!}
\end{aligned}$$

This limit is the slope of the curve at the y-intercept, i.e., $f'(0)$, if it exists.

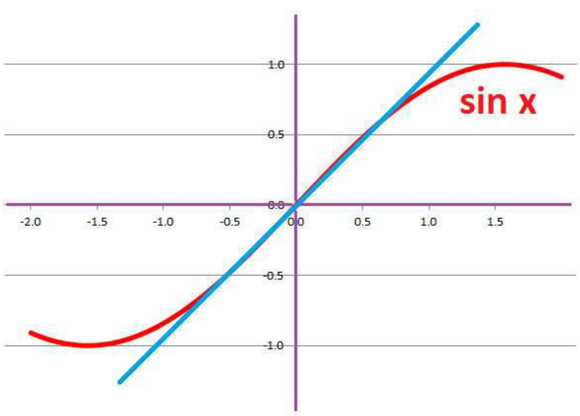


Indeed, the natural base exponential function is a special one!

We would like to find the derivatives of *sine and cosine* next.

Example 3.9.9: derivative of $\sin$ at 0

Before addressing the general case, let's recall their derivatives at $x = 0$. We know that the graph of $y = \sin x$ crosses the y-axis at 45 degrees:



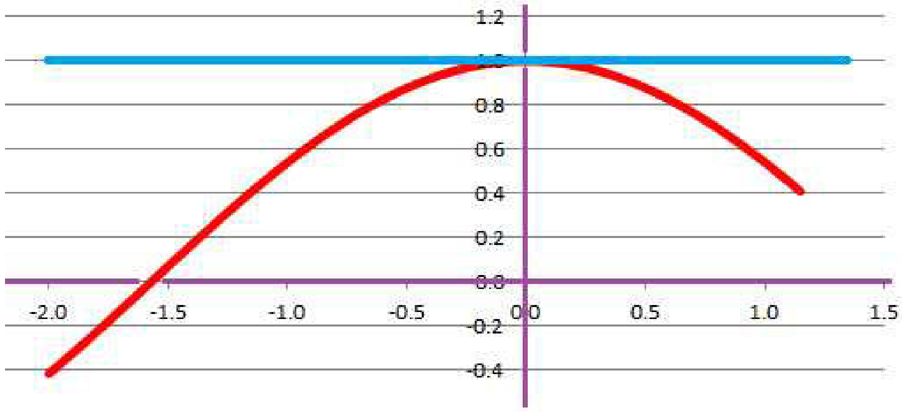
This means that the following famous trigonometric limit,

$$\lim_{x\to 0}\frac{\sin x}{x}=1\,,$$

has now a new interpretation:

$$(\sin x)'\Big|_{x=0}=1\,.$$

We also know that the graph of $y=\cos x$ crosses the y-axis horizontally:



This means that the following famous trigonometric limit,

$$\lim_{x\to 0}\frac{1-\cos x}{x}=0\,,$$

has now a new interpretation:

$$(\cos x)'\Big|_{x=0}=0\,.$$

We now use these results to find the derivative functions of sine and cosine. Remarkably, the derivative of one is the other, up to a sign.

Theorem 3.9.10: Difference of Sine and Cosine

Suppose we have a mid-point partition; i.e., the nodes are $x=a,\ a+h$ and the secondary node is $c=a+h/2$. Then the differences of $y=\sin x$ and $y=\cos x$ are given at c by the following:

$$\Delta(\sin x)=2\sin(h/2)\cdot\ \cos c\quad\text{and}\quad\Delta(\cos x)=-2\sin(h/2)\cdot\ \sin c\,.$$

Proof.

First $f(x)=\sin x$. We use the following formula (seen in Volume 1, Chapter 1PC-4):

$$\sin u-\sin v=2\sin\frac{u-v}{2}\cos\frac{u+v}{2}\,.$$

This formula is the reason why we choose this particular partition. We compute at c:

$$\begin{aligned}\Delta f\ &=\sin(a+h)-\sin(a)\\ &=2\sin(h/2)\cos(a+h/2)\,.\qquad\text{Here } a+h/2 \text{ is exactly the secondary node.}\end{aligned}$$

The proof of the second identity uses another trig formula:

$$\cos u-\cos v=-2\sin\frac{u-v}{2}\sin\frac{u+v}{2}\,.$$

Theorem 3.9.11: Difference Quotient of Sine and Cosine

Suppose we have a mid-point partition; i.e., the nodes are $x = a,\ a + h$ and the secondary node is $c = a + h/2$. Then the difference quotients of $y = \sin x$ and $y = \cos x$ are given at c by the following:

$$\frac{\Delta}{\Delta x}(\sin x) = \frac{\sin(h/2)}{h/2} \cdot \quad \cos c \quad \textit{and} \quad \frac{\Delta}{\Delta x}(\cos x) = -\frac{\sin(h/2)}{h/2} \cdot \quad \sin c\,.$$

Proof.

First $f(x) = \sin x$. We compute at c from the last theorem:

$$\begin{aligned} \frac{\Delta f}{\Delta x} &= \frac{2\sin(h/2)\cos(a + h/2)}{h} \\ &= \frac{\sin(h/2)}{h/2} \cdot \cos(a + h/2)\,. \end{aligned}$$

Theorem 3.9.12: Derivative of Sine and Cosine

The derivatives of $y = \sin x$ and $y = \cos x$ are given by:

$$\boxed{\frac{d}{dx}(\sin x) = \quad \cos x \quad \textit{and} \quad \frac{d}{dx}(\cos x) = - \quad \sin x}$$

Proof.

First $f(x) = \sin x$. We have at c:

$$\begin{array}{lll} \dfrac{\Delta f}{\Delta x} = \dfrac{\sin(h/2)}{h/2} & \cdot \cos(a + h/2) & \\ \to 1 & \cdot \cos a & \text{as } \ h \to 0 \\ = & \cos a\,. & \end{array}$$

The last step is the first famous limit above combined with the continuity of $\cos x$.

The proof of the second identity uses the other famous limit above.

Exercise 3.9.13

Provide a proof of the missing part.

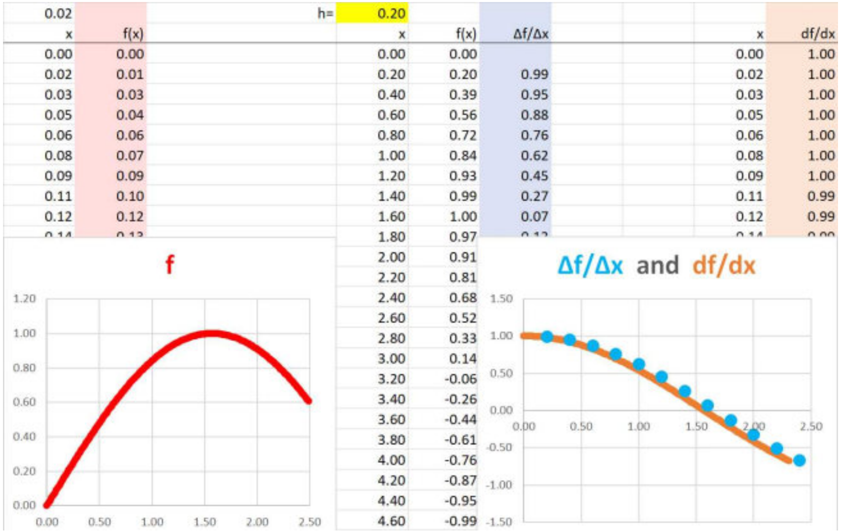
Let's compare the formulas for the difference quotient of $\sin x$ with its derivative:

$$\begin{array}{llll} \text{(A)} & \dfrac{\Delta}{\Delta x}(\sin x) & = \dfrac{\sin(h/2)}{h/2} \cdot & \cos(x + h/2)\,. \\ \text{(B)} & \dfrac{d}{dx}(\sin x) & = & \cos x\,. \end{array}$$

We know that

$$0 < \frac{\sin(h/2)}{h/2} < 1\,,$$

for a small enough h. Therefore, the graph of the difference quotient differs from that of the derivative by a vertical shrink and, in addition, a horizontal shift. As h is approaching 0, the effect of these two operations diminishes; the result is the *convergence* of the graph of the former to the graph of the latter:



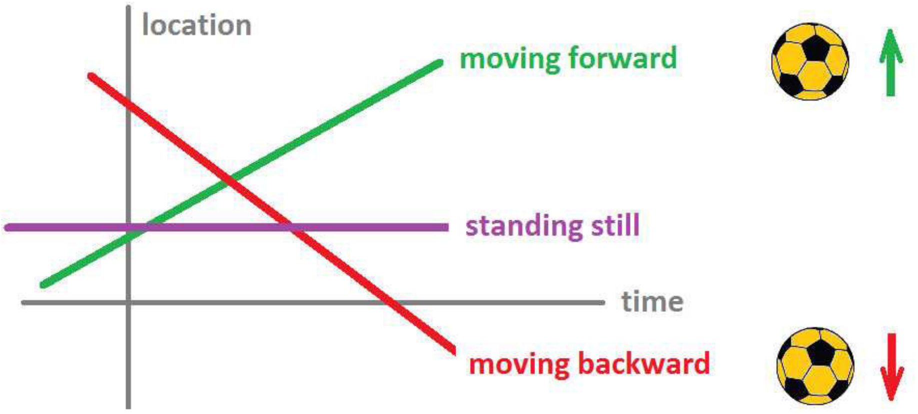
The formulas in the Lagrange notation are as follows:

$$\boxed{(\sin x)' = \cos x \quad \text{and} \quad (\cos x)' = -\sin x}$$

3.10. Free fall

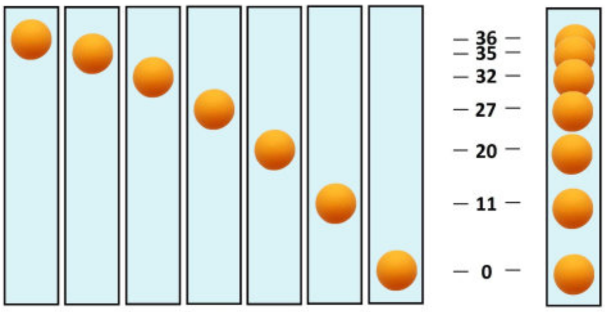
Example 3.10.1: moving ball

We know that a *ball rolling* on a horizontal plane will have a constant velocity:



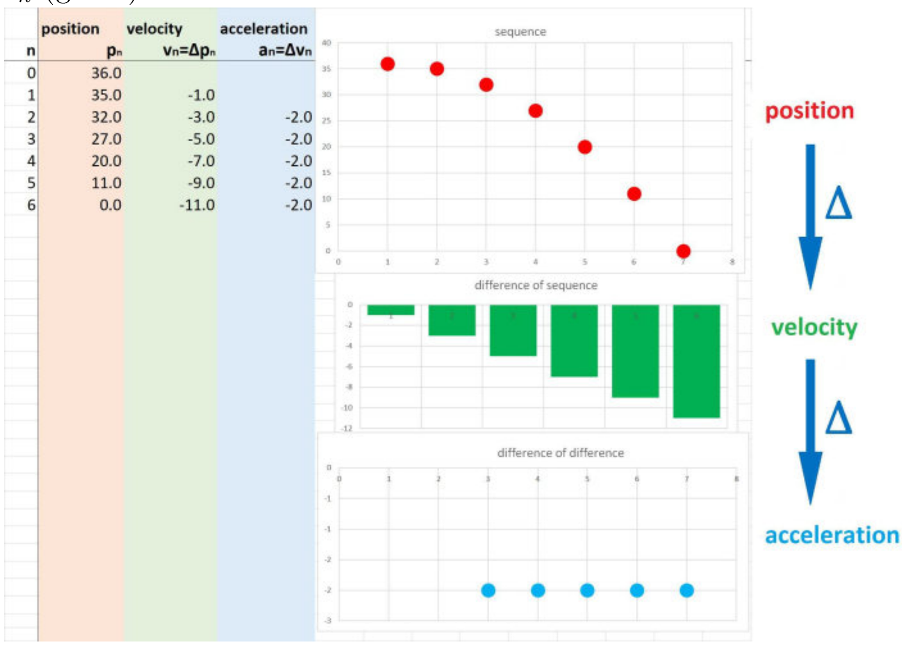
What if the ball is now thrown *up in the air*? The dynamics is very different. In the former, as there is no force changing the velocity, the latter remains constant. In the latter, the velocity is constantly changed by the gravity.

Imagine that we have this experimental data of the heights of a ping-pong ball falling down recorded about every .1 second, measured in inches:



We use a spreadsheet to plot the *location* sequence, p_n (red). We then compute the difference of p_n,

i.e., the *velocity*, v_n (green):



n	position p_n	velocity $v_n=\Delta p_n$	acceleration $a_n=\Delta v_n$
0	36.0		
1	35.0	-1.0	
2	32.0	-3.0	-2.0
3	27.0	-5.0	-2.0
4	20.0	-7.0	-2.0
5	11.0	-9.0	-2.0
6	0.0	-11.0	-2.0

It looks like a straight line. But this time, we take one more step: We compute the difference of the velocity sequence. It is the *acceleration*, a_n (blue). It appears constant! There might be a law of nature here.

Let's accept the premise we've put forward:

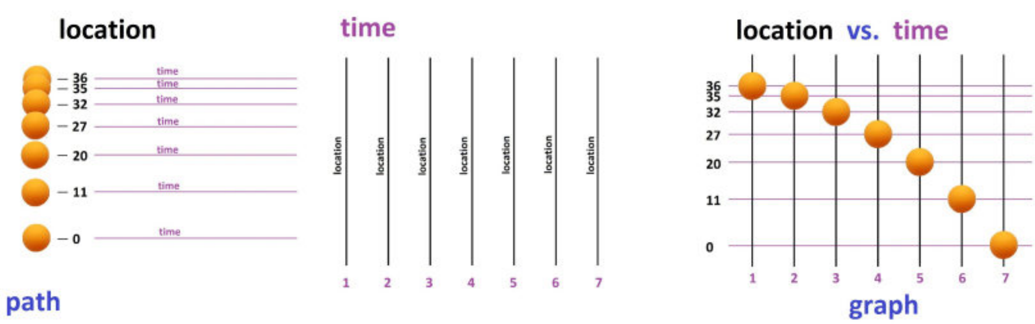
▶ *The acceleration of free fall is constant.*

Then we can try to predict the behavior of an object thrown in the air – from any initial height and with any initial velocity. The direction of our computation is opposite to that of the last example:

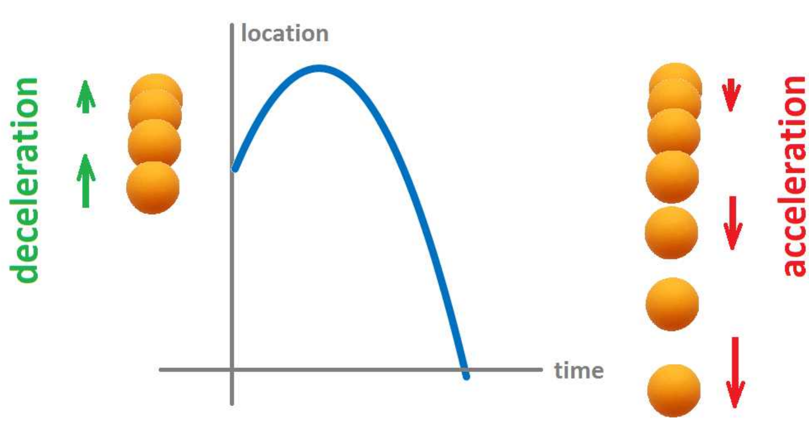
▶ We use our knowledge of the acceleration to derive the velocity, and then derive the position of the object in time.

While we used *differences* in the last example, we use *sums* (seen in Volume 1, Chapter 1PC-1) now.

We plot these positions against time:



The graph "looks like" a parabola:



Earlier, we used these difference quotient formulas to find the velocity from the position and then the acceleration from the velocity:

$$v_n = \frac{\Delta p}{\Delta t} = \frac{p_{n+1} - p_n}{h} \quad \text{and} \quad a_n = \frac{\Delta v}{\Delta t} = \frac{v_{n+1} - v_n}{h},$$

where h is the increment of time. The dependence of the velocity on the position and of the acceleration on the velocity is, of course, identical.

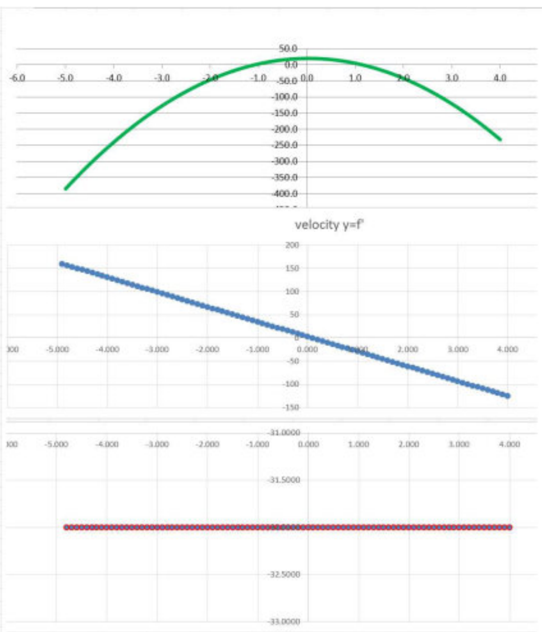
Example 3.10.2: free fall

The following formula is used again:

```
=(RC[-1]-R[-1]C[-1])/R2C1
```

These are the results:

time	location	-->	time	velocity	-->	time	accel
t	y=f(t)		t	y=f'		t	f''
-5.000	-385.000		-5.000			-5.000	
-4.900	-369.060		-4.900	159.4		-4.900	
-4.800	-353.440		-4.800	156.2		-4.800	-32.0000
-4.700	-338.140		-4.700	153		-4.700	-32.0000
-4.600	-323.160		-4.600	149.8		-4.600	-32.0000
-4.500	-308.500		-4.500	146.6		-4.500	-32.0000
-4.400	-294.160		-4.400	143.4		-4.400	-32.0000
-4.300	-280.140		-4.300	140.2		-4.300	-32.0000
-4.200	-266.440		-4.200	137		-4.200	-32.0000
-4.100	-253.060		-4.100	133.8		-4.100	-32.0000
-4.000	-240.000		-4.000	130.6		-4.000	-32.0000
-3.900	-227.260		-3.900	127.4		-3.900	-32.0000
-3.800	-214.840		-3.800	124.2		-3.800	-32.0000
-3.700	-202.740		-3.700	121		-3.700	-32.0000
-3.600	-190.960		-3.600	117.8		-3.600	-32.0000
-3.500	-179.500		-3.500	114.6		-3.500	-32.0000
-3.400	-168.360		-3.400	111.4		-3.400	-32.0000
-3.300	-157.540		-3.300	108.2		-3.300	-32.0000
-3.200	-147.040		-3.200	105		-3.200	-32.0000
-3.100	-136.860		-3.100	101.8		-3.100	-32.0000
-3.000	-127.000		-3.000	98.6		-3.000	-32.0000
-2.900	-117.460		-2.900	95.4		-2.900	-32.0000
-2.800	-108.240		-2.800	92.2		-2.800	-32.0000
-2.700	-99.340		-2.700	89		-2.700	-32.0000
-2.600	-90.760		-2.600	85.8		-2.600	-32.0000
-2.500	-82.500		-2.500	82.6		-2.500	-32.0000
-2.400	-74.560		-2.400	79.4		-2.400	-32.0000
-2.300	-66.940		-2.300	76.2		-2.300	-32.0000
-2.200	-59.640		-2.200	73		-2.200	-32.0000
-2.100	-52.660		-2.100	69.8		-2.100	-32.0000
-2.000	-46.000		-2.000	66.6		-2.000	-32.0000
-1.900	-39.660		-1.900	63.4		-1.900	-32.0000
-1.800	-33.640		-1.800	60.2		-1.800	-32.0000
-1.700	-27.940		-1.700	57		-1.700	-32.0000
-1.600	-22.560		-1.600	53.8		-1.600	-32.0000
-1.500	-17.500		-1.500	50.6		-1.500	-32.0000
-1.400	-12.760		-1.400	47.4		-1.400	-32.0000
-1.300	-8.340		-1.300	44.2		-1.300	-32.0000
-1.200	-4.240		-1.200	41		-1.200	-32.0000
-1.100	-0.460		-1.100	37.8		-1.100	-32.0000
-1.000	3.000		-1.000	34.6		-1.000	-32.0000
-0.900	6.140		-0.900	31.4		-0.900	-32.0000
-0.800	8.960		-0.800	28.2		-0.800	-32.0000
-0.700	11.460		-0.700	25		-0.700	-32.0000
-0.600	13.640		-0.600	21.8		-0.600	-32.0000
-0.500	15.500		-0.500	18.6		-0.500	-32.0000
-0.400	17.040		-0.400	15.4		-0.400	-32.0000
-0.300	18.260		-0.300	12.2		-0.300	-32.0000



Now in reverse!

To create formulas for a simulation of free fall, the derivation goes in the opposite direction:

- the velocity from the acceleration, and then
- the location from the acceleration.

The formulas are solved for p_{n+1} and v_{n+1} respectively:

$$\begin{aligned} v_n &= \frac{p_{n+1} - p_n}{h} &\implies& \quad v_{n+1} = v_n + ha_n \\ a_n &= \frac{v_{n+1} - v_n}{h} &\implies& \quad p_{n+1} = p_n + hv_n \end{aligned}$$

The dependence of the velocity on the acceleration and of the position on the velocity is, of course, identical.

Warning!

Unlike the former, these are *recursive* sequences.

The formulas are just versions of the same elementary school formula:

$$\text{speed} = \text{distance} \,/\, \text{time} \qquad \text{and} \qquad \text{distance} = \text{speed} \times \text{time}$$

Summary:

		vertical
position		p_n given
velocity		$v_n = \dfrac{p_{n+1} - p_n}{h}$
acceleration		$a_n = \dfrac{v_{n+1} - v_n}{h}$

$\Longrightarrow$

		vertical
acceleration		a_n given
velocity		$v_{n+1} = v_n + ha_n$
position		$p_{n+1} = p_n + hv_n$

Warning!

We leave the first column empty to be used later to record the horizontal progress when the ball is thrown forward.

Example 3.10.3: free fall

Let's consider a specific problem.

▶ *PROBLEM:* From a 100-foot building, a ball is thrown up at 50 feet per second in such a way that it falls on the ground. How high will the ball go?

We use the same spreadsheet formula for the velocity and position:

```
=R[-1]C+R[-1]C[-1]*R2C1
```

Now in the specific case of *free fall*, there is just one force, the gravity, and the vertical acceleration is known to be $a = -g$, where g is the gravitational constant:

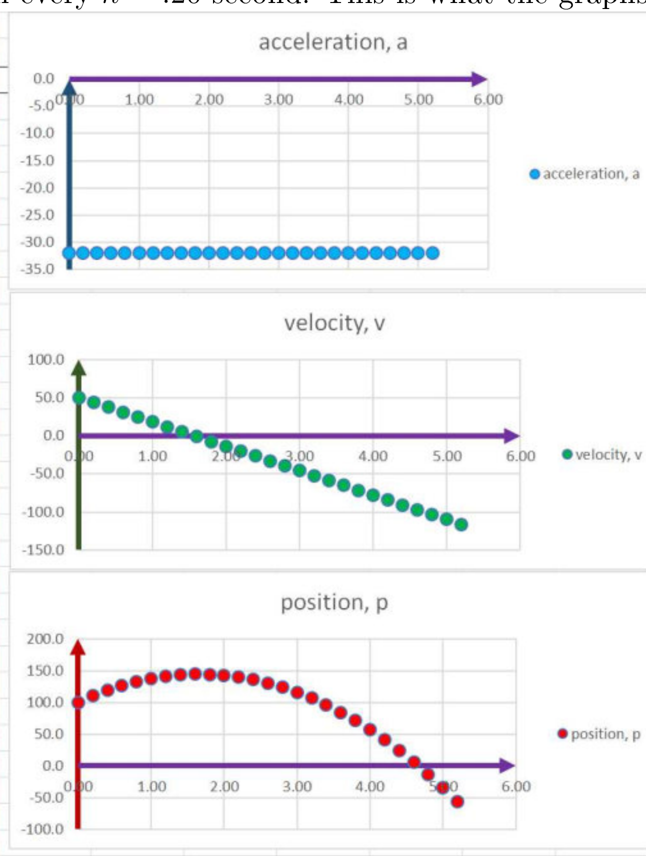
$$g = 32 \text{ ft/sec}^2 .$$

Next, we acquire the initial conditions:

- The initial location is given by: $p_0 = 100$.
- The initial velocity is given by: $v_0 = 50$.

We use the formulas to evaluate the location every $h = .20$ second. This is what the graphs look like:

h				
0.20	t	a	v	p
intial:	0.00	-32.00	50.00	100.00
	0.20	-32.00	43.60	110.00
	0.40	-32.00	37.20	118.72
	0.60	-32.00	30.80	126.16
	0.80	-32.00	24.40	132.32
	1.00	-32.00	18.00	137.20
	1.20	-32.00	11.60	140.80
	1.40	-32.00	5.20	143.12
	1.60	-32.00	-1.20	144.16
	1.80	-32.00	-7.60	143.92
	2.00	-32.00	-14.00	142.40
	2.20	-32.00	-20.40	139.60
	2.40	-32.00	-26.80	135.52
	2.60	-32.00	-33.20	130.16
	2.80	-32.00	-39.60	123.52
	3.00	-32.00	-46.00	115.60
	3.20	-32.00	-52.40	106.40
	3.40	-32.00	-58.80	95.92
	3.60	-32.00	-65.20	84.16
	3.80	-32.00	-71.60	71.12
	4.00	-32.00	-78.00	56.80
	4.20	-32.00	-84.40	41.20
	4.40	-32.00	-90.80	24.32
	4.60	-32.00	-97.20	6.16
	4.80	-32.00	-103.60	-13.28
	5.00	-32.00	-110.00	-34.00
	5.20	-32.00	-116.40	-56.00
	5.40	-32.00	-122.80	-79.28
	5.60	-32.00	-129.20	-103.84
	5.80	-32.00	-135.60	-129.68



By simply examining the data, we can solve various problems about this experiment:

1. To find the highest elevation, we look at the row with p. The largest value seems to be close to $y = 144$ feet.

2. To find when the ball hits the ground, we scroll down to find the row with p close to 0. It happens sometime close to $t = 4.7$ seconds.
3. To find how fast the ball hits the ground, we scroll down again to find the row with p close to 0 and look up the value of v. It is close to $v = 200$ feet per second.

Of course, we can decrease the time increment $h = \Delta x$ and get more accurate answers.

With the spreadsheet, we can ask and answer a variety of other questions about such motion (how hard it hits the ground, etc.). However, we can only do one example at a time! The conclusions we draw are specific to these initial conditions (as well as gravity etc.). The results are also dependent on Δx. This why we now proceed to the *continuous* case.

We take the limit:

$$h = \Delta x \to 0\,.$$

This time, instead of sequences, we have these *functions of time*:

- p is the height, the vertical location.
- $v = p'$ is the vertical velocity.
- $a = v'$ is the vertical acceleration.

Now the specific case of free fall:

$$a = -g\,.$$

We have conjectured in this chapter that:

- The derivative of a quadratic polynomial is linear.
- The derivative of a linear polynomial is constant.

We will show in Chapter 5 that, conversely:

- The only function the derivative of which is linear is a quadratic polynomial.
- The only function the derivative of which is constant is a linear polynomial.

From the latter two, we derive:

$$a = a(t) \text{ is constant } \implies v = v(t) \text{ is linear} \implies p = p(t) \text{ is quadratic.}$$

In other words, we have:

$$p(t) = ax^2 + bx + c\,.$$

What makes these specific are the *initial conditions*:

- p_0 is the initial height, $p_0 = p(0)$.
- v_0 is the initial vertical component of velocity, $v(0) = \left.\dfrac{dp}{dt}\right|_{t=0}$.

Therefore, we have:

$$\boxed{p(t) \quad = p_0 \quad + v_0 t \quad -\frac{1}{2} g t^2}$$

Example 3.10.4: free fall

In the problem of ours, we have:

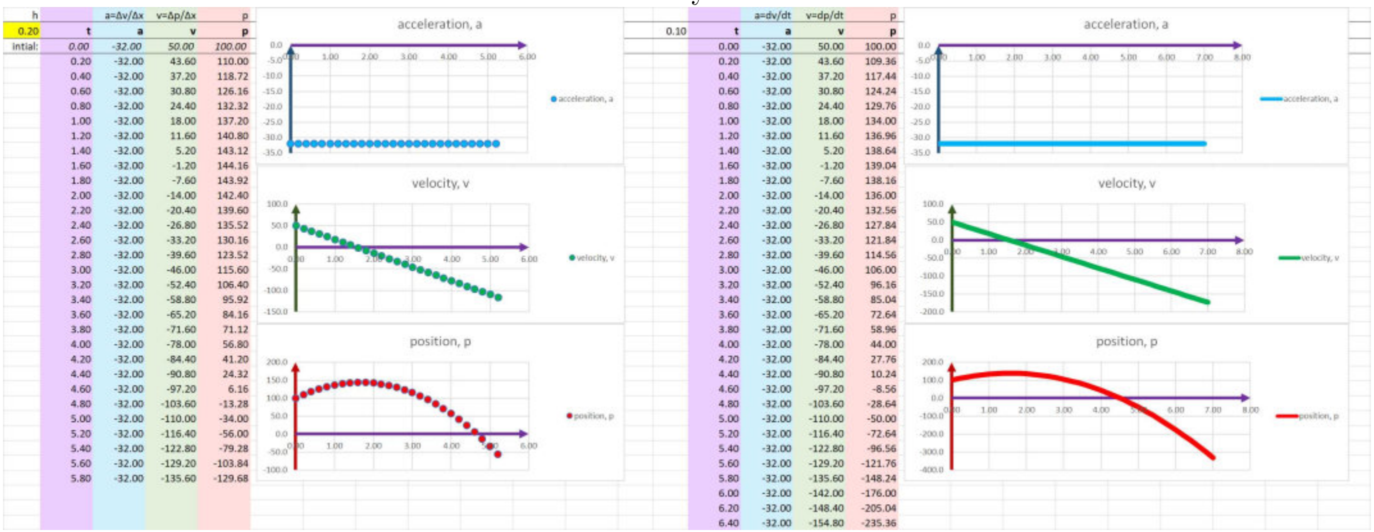
$$p_0 = 100,\ v_0 = 50\,.$$

Our equation becomes:

$$p \quad = 100 \quad + 50t \quad - 16t^2\,.$$

In contrast to the discrete case, the formula for the position isn't recursive but direct and explicit!

Before we utilize the explicit algebraic representation, we visualize the results by plotting the graph of this function next to the one obtained recursively:



h		a=Δv/Δx	v=Δp/Δx	p
0.20	t	a	v	p
intial:	*0.00*	*-32.00*	*50.00*	*100.00*
	0.20	-32.00	43.60	110.00
	0.40	-32.00	37.20	118.72
	0.60	-32.00	30.80	126.16
	0.80	-32.00	24.40	132.32
	1.00	-32.00	18.00	137.20
	1.20	-32.00	11.60	140.80
	1.40	-32.00	5.20	143.12
	1.60	-32.00	-1.20	144.16
	1.80	-32.00	-7.60	143.92
	2.00	-32.00	-14.00	142.40
	2.20	-32.00	-20.40	139.60
	2.40	-32.00	-26.80	135.52
	2.60	-32.00	-33.20	130.16
	2.80	-32.00	-39.60	123.52
	3.00	-32.00	-46.00	115.60
	3.20	-32.00	-52.40	106.40
	3.40	-32.00	-58.80	95.92
	3.60	-32.00	-65.20	84.16
	3.80	-32.00	-71.60	71.12
	4.00	-32.00	-78.00	56.80
	4.20	-32.00	-84.40	41.20
	4.40	-32.00	-90.80	24.32
	4.60	-32.00	-97.20	6.16
	4.80	-32.00	-103.60	-13.28
	5.00	-32.00	-110.00	-34.00
	5.20	-32.00	-116.40	-56.00
	5.40	-32.00	-122.80	-79.28
	5.60	-32.00	-129.20	-103.84
	5.80	-32.00	-135.60	-129.68

		a=dv/dt	v=dp/dt	p
0.10	t	a	v	p
	0.00	-32.00	50.00	100.00
	0.20	-32.00	43.60	109.36
	0.40	-32.00	37.20	117.44
	0.60	-32.00	30.80	124.24
	0.80	-32.00	24.40	129.76
	1.00	-32.00	18.00	134.00
	1.20	-32.00	11.60	136.96
	1.40	-32.00	5.20	138.64
	1.60	-32.00	-1.20	139.04
	1.80	-32.00	-7.60	138.16
	2.00	-32.00	-14.00	136.00
	2.20	-32.00	-20.40	132.56
	2.40	-32.00	-26.80	127.84
	2.60	-32.00	-33.20	121.84
	2.80	-32.00	-39.60	114.56
	3.00	-32.00	-46.00	106.00
	3.20	-32.00	-52.40	96.16
	3.40	-32.00	-58.80	85.04
	3.60	-32.00	-65.20	72.64
	3.80	-32.00	-71.60	58.96
	4.00	-32.00	-78.00	44.00
	4.20	-32.00	-84.40	27.76
	4.40	-32.00	-90.80	10.24
	4.60	-32.00	-97.20	-8.56
	4.80	-32.00	-103.60	-28.64
	5.00	-32.00	-110.00	-50.00
	5.20	-32.00	-116.40	-72.64
	5.40	-32.00	-122.80	-96.56
	5.60	-32.00	-129.20	-121.76
	5.80	-32.00	-135.60	-148.24
	6.00	-32.00	-142.00	-176.00
	6.20	-32.00	-148.40	-205.04
	6.40	-32.00	-154.80	-235.36

Just as in the last example, we can use this plot to solve a variety of problems about this motion. Let's revisit the two problems about this specific throw we solved numerically. They are solved the same way:

1. To find the highest elevation, we look at the row with p. The largest value seems to be close to $y = 139$ feet.
2. To find when the ball hits the ground, we scroll down to find the row with p close to 0. It happens sometime close to $t = 4.5$ seconds.

These are just approximations. The formula, however, gives us a way to answer the questions with absolute accuracy. We can even avoid differentiation.

For the first problem, we realize that $p = -16t^2 + 50t + 100$ is a parabola! And the vertex of $y = ax^2 + bx + c$ is at $x = -b/a$ (Volume 1, Chapter 1PC-4). Therefore, the highest point is reached at time

$$t = -50/(-2 \cdot 16) = 1.5625\,.$$

Then, the highest elevation is

$$p = 100 + 50 \cdot 1.5625 - 16 \cdot 1.5625^2 = 139.0625\,.$$

The result matches our estimate!

For the second problem, the altitude at the end is 0, so to find *when* it happened, we set $p = 0$, or

$$-16t^2 + 50t + 100 = 0\,,$$

and solve for t. According to the *Quadratic Formula* (seen in Volume 1, Chapter 1PC-4), we have:

$$t = \frac{-50 - \sqrt{50^2 - 4(-16)100}}{2 \cdot (-16)} \approx 4.5106\,.$$

The result matches our estimate!

Exercise 3.10.5

What happened to $\pm$?

Exercise 3.10.6

How high does the projectile go in the above example?

Exercise 3.10.7

Using the above example, how long will it take for the projectile to reach the ground if fired *down*?

Exercise 3.10.8

Use the above model to determine how long it will take for an object to reach the ground if it is dropped. Make up your own questions about the situation and answer them. Repeat.

Example 3.10.9: acceleration that depends on time

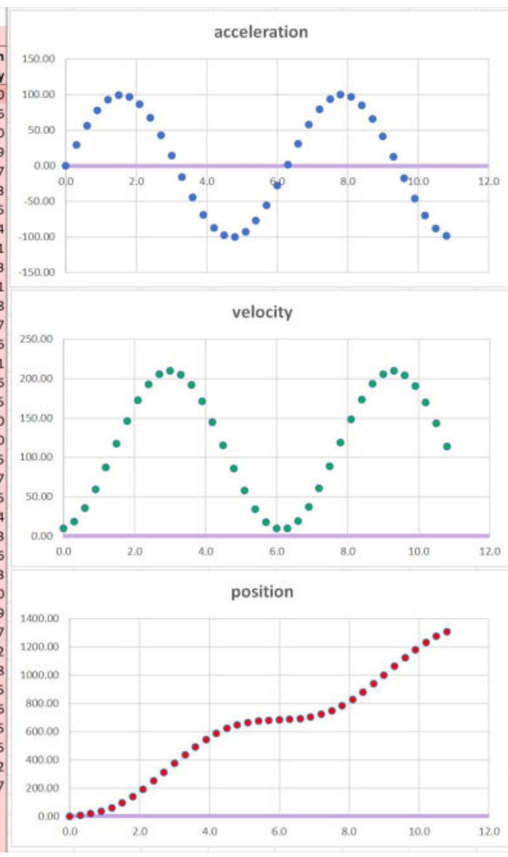
What your algebraic analysis is incapable of doing (for now) is to handle the case of time-dependent acceleration. For example, imagine that there is another planet that moves in and out periodically. Then the acceleration will be also changing periodically. What will be the motion of the same ball?

The spreadsheet works with no change! We just insert a formula for the acceleration, for example:

```
=100*SIN(RC[-1])
```

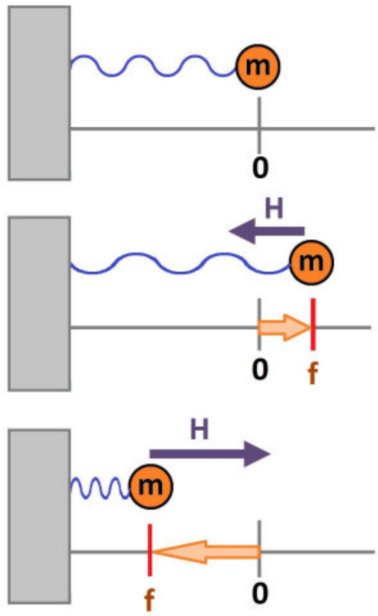
Then the rest is taken care of:

h			
0.30	"=100*SIN(RC[-1])"	"=R[-1]C+RC[-1]*R2C1"	"=R[-1]C+RC[-1]*R2C1"
time	acceleration	velocity	position
t	a	v	y
0.00	0.00	10.00	0.00
0.30	29.55	18.87	5.66
0.60	56.46	35.80	16.40
0.90	78.33	59.30	34.19
1.20	93.20	87.27	60.37
1.50	99.75	117.19	95.53
1.80	97.38	146.41	139.45
2.10	86.32	172.30	191.14
2.40	67.55	192.57	248.91
2.70	42.74	205.39	310.53
3.00	14.11	209.62	373.41
3.30	-15.77	204.89	434.88
3.60	-44.25	191.61	492.37
3.90	-68.78	170.98	543.66
4.20	-87.16	144.83	587.11
4.50	-97.75	115.51	621.76
4.80	-99.62	85.62	647.45
5.10	-92.58	57.85	664.80
5.40	-77.28	34.66	675.20
5.70	-55.07	18.14	680.65
6.00	-27.94	9.76	683.57
6.30	1.68	10.27	686.65
6.60	31.15	19.61	692.54
6.90	57.84	36.97	703.63
7.20	79.37	60.78	721.86
7.50	93.80	88.92	748.53
7.80	99.85	118.87	784.20
8.10	96.99	147.97	828.59
8.40	85.46	173.61	880.67
8.70	66.30	193.50	938.72
9.00	41.21	205.86	1000.48
9.30	12.45	209.59	1063.35
9.60	-17.43	204.36	1124.66
9.90	-45.75	190.64	1181.85
10.20	-69.99	169.64	1232.75
10.50	-87.97	143.25	1275.72
10.80	-98.09	113.82	1309.87



Example 3.10.10: acceleration that depends on location

On the next level is the case when the acceleration is not known as a function of time! We just know how it depends on location. For example, the force of the spring is proportional to the negative of the displacement from the equilibrium.

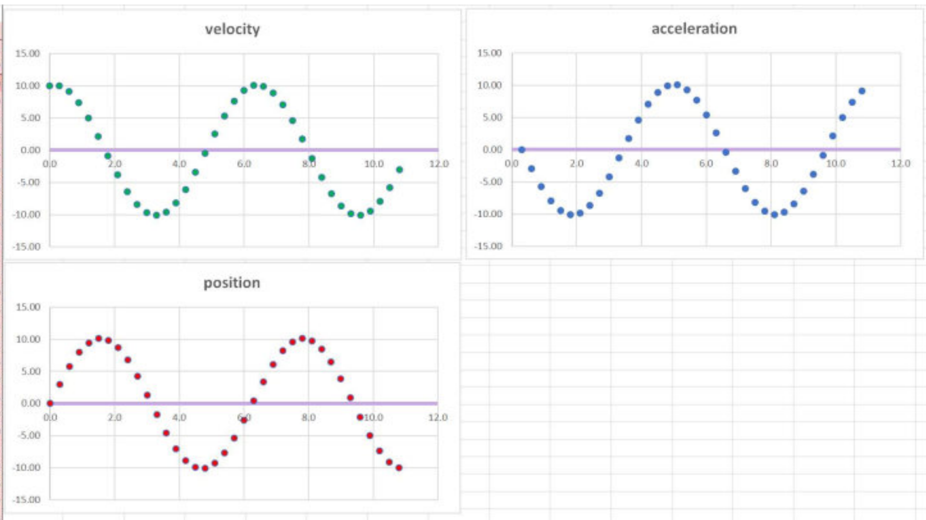


We modify the spreadsheet by giving the acceleration a formula that makes a reference to the location:

```
=-R[-1]C[2]
```

The result is oscillation:

h			
0.30	"=-R[-1]C[2]"	"=R[-1]C+RC[-1]*R2C1"	"=R[-1]C+RC[-1]*R2C1"
time t	acceleration a	velocity v	position y
0.00		10.00	0.00
0.30	0.00	10.00	3.00
0.60	-3.00	9.10	5.73
0.90	-5.73	7.38	7.94
1.20	-7.94	5.00	9.44
1.50	-9.44	2.16	10.09
1.80	-10.09	-0.86	9.83
2.10	-9.83	-3.81	8.69
2.40	-8.69	-6.42	6.76
2.70	-6.76	-8.45	4.23
3.00	-4.23	-9.72	1.31
3.30	-1.31	-10.11	-1.72
3.60	1.72	-9.60	-4.60
3.90	4.60	-8.22	-7.06
4.20	7.06	-6.10	-8.89
4.50	8.89	-3.43	-9.92
4.80	9.92	-0.45	-10.06
5.10	10.06	2.56	-9.29
5.40	9.29	5.35	-7.68
5.70	7.68	7.66	-5.39
6.00	5.39	9.27	-2.60
6.30	2.60	10.05	0.41
6.60	-0.41	9.93	3.39
6.90	-3.39	8.91	6.06
7.20	-6.06	7.09	8.19
7.50	-8.19	4.64	9.58

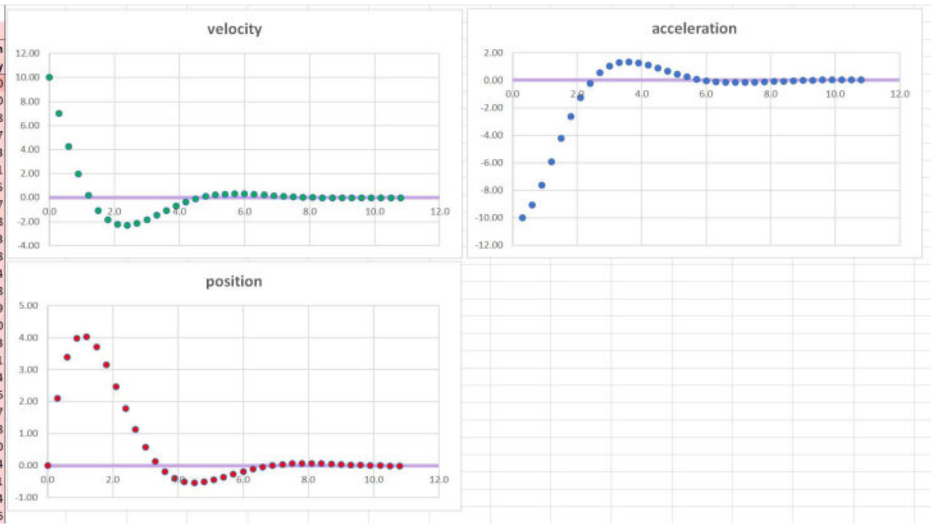


The acceleration can also depend on the velocity! For example, it may be proportional to its negative as in the case of air resistance. We just modify the spreadsheet by giving the acceleration a formula that makes a reference to both the location and the velocity:

```
=-R[-1]C[2]-R[-1]C[1]
```

The result is a quick disappearance of the oscillation:

h			
0.30	"=-R[-1]C[2]-R[-1]C[1]"	"=R[-1]C+RC[-1]*R2C1"	"=R[-1]C+RC[-1]*R2C1"
time t	acceleration a	velocity v	position y
0.00		10.00	0.00
0.30	-10.00	7.00	2.10
0.60	-9.10	4.27	3.38
0.90	-7.65	1.97	3.97
1.20	-5.95	0.19	4.03
1.50	-4.22	-1.08	3.71
1.80	-2.63	-1.87	3.15
2.10	-1.28	-2.25	2.47
2.40	-0.22	-2.32	1.78
2.70	0.54	-2.16	1.13
3.00	1.02	-1.85	0.58
3.30	1.27	-1.47	0.14
3.60	1.33	-1.07	-0.18
3.90	1.25	-0.69	-0.39
4.20	1.08	-0.37	-0.50
4.50	0.87	-0.11	-0.53
4.80	0.64	0.09	-0.51
5.10	0.42	0.21	-0.44
5.40	0.23	0.28	-0.36
5.70	0.08	0.31	-0.27
6.00	-0.04	0.29	-0.18
6.30	-0.11	0.26	-0.10
6.60	-0.16	0.21	-0.04
6.90	-0.17	0.16	0.01
7.20	-0.17	0.11	0.04
7.50	-0.15	0.06	0.06



Chapter 4: Differentiation

Contents

4.1. Differentiation over addition and constant multiple: linearity

In this chapter, we will be taking a broader look at how we compute the rate of change.

If a function f is defined at the nodes $x_0, x_1, ..., x_n$ of a partition of interval $[a,b]$, it is simply a sequence of numbers. And so is its *difference*:

$$\Delta f = f(x_{n+1}) - f(x_n)\,.$$

For example:

$$\Delta(x^2) = x_{n+1}^2 - x_n^2\,.$$

What this means is that this procedure is a special kind of function, a *function of functions*:

$$\text{function} \quad \to \quad \boxed{\quad \Delta \quad} \quad \to \quad \text{another function}$$

To find the *difference quotient* $\dfrac{\Delta f}{\Delta x}$ of f, we just divide the difference by $\Delta x = x_{n+1} - x_n$. We have, for example:

$$\frac{\Delta(x^2)}{\Delta x} = \frac{x_{n+1}^2 - x_n^2}{\Delta x}\,.$$

We have created another function of functions:

$$\text{function} \quad \to \quad \boxed{\;\frac{\Delta}{\Delta x}\;} \quad \to \quad \text{another function}$$

Next, the *derivative* is defined as a limit. It is the limit of the difference quotient. Therefore, unlike the limits we saw prior to the derivatives, this one has a parameter, the location x. We have, for example:

$$\frac{d(x^2)}{dx} = \lim_{\Delta x \to 0} \frac{x_{n+1}^2 - x_n^2}{\Delta x}.$$

That is why if the input of this limit is a differentiable function f, then the output is another function f'. What this means is that this process is a special kind of function too, a function of functions:

$$\text{function} \quad \to \quad \boxed{\frac{d}{dx}} \quad \to \quad \text{another function}$$

We need to understand how these three functions operate. We will develop shortcuts and algebraic rules for evaluating differences, difference quotients, and derivatives.

The last one can be found without resorting to using limits! Here is an example from the last chapter:

$$ax^2 + bx + c \quad \to \quad \boxed{\frac{d}{dx}} \quad \to \quad 2ax + b$$

The question we will be asking is the following:

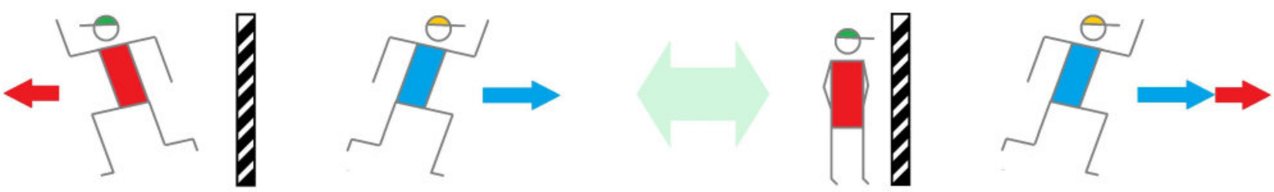
▶ What happens to the output function of differentiation as we perform algebraic operations with the input functions?

There are a few shortcut properties.

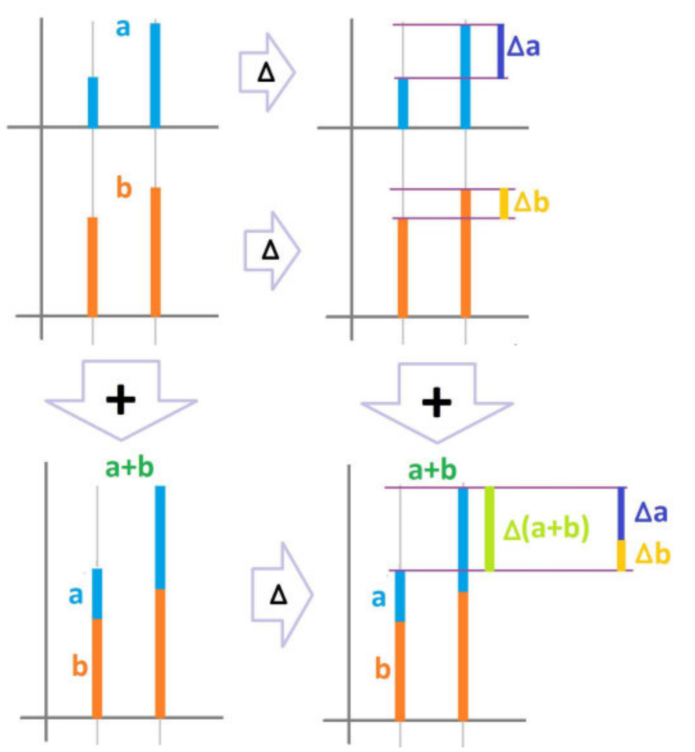
Let's take another look at this elementary statement about *motion*:

▶ `IF` two runners are running away from a post, `THEN` their relative velocity is the sum of their respective velocities.

It's as if the one runner is standing still while the other is running with the combined speed:



The idea why we *add* their differences when we add functions is illustrated below:



Here, either of the two routes through the diagram produces the same result:

- Down, then right: First the bars that represent the values of the sequences are stacked on top of each other, then the resulting difference of heights is found.
- Right, then down: First differences of the bars that represent the values of either sequence are found and then stacked on top of each other.

The algebra behind this geometry is very simple:

$$(A+B)-(a+b)=(A-a)+(B-b)\,.$$

It's the *Associative Rule* of addition.

This is the conclusion:

Theorem 4.1.1: Sum Rule for Differences

The difference of the sum of two functions is the sum of their differences.

In other words, for any two functions f,g, their differences satisfy:

$$\boxed{\Delta(f+g)=\Delta f+\Delta g}$$

Proof.

Applying the definition to the function $f+g$, we have:

$$\begin{aligned}\Delta(f+g)(c)\ &=(f+g)(x+\Delta x)-(f+g)(x)\\ &=f(x+\Delta x)+g(x+\Delta x)-f(x)-g(x)\\ &=\big(f(x+\Delta x)-f(x)\big)+\big(g(x+\Delta x)-g(x)\big)\\ &=\Delta f(c)+\Delta g(c)\,.\end{aligned}$$

Simply dividing this formula by Δx will produce its analog for difference quotients:

Theorem 4.1.2: Sum Rule for Difference Quotients

The difference quotient of the sum of two functions is the sum of their difference quotients.

In other words, for any two functions f,g defined at the adjacent nodes x and $x+\Delta x$ of a partition, the difference quotients (defined at the corresponding secondary node) satisfy:

$$\boxed{\frac{\Delta(f+g)}{\Delta x}=\frac{\Delta f}{\Delta x}+\frac{\Delta g}{\Delta x}}$$

The limit of the latter as $\Delta x\to 0$ will produce the analog for derivatives:

Theorem 4.1.3: Sum Rule for Derivatives

The sum of two functions differentiable at a point is differentiable at that point and its derivative is equal to the sum of their derivatives.

In other words, for any two functions f, g differentiable at x, we have at x:

$$\boxed{\frac{d(f+g)}{dx} = \frac{df}{dx} + \frac{dg}{dx}}$$

Proof.

The limit with $c = x$:

$$\begin{aligned} \frac{\Delta(f+g)}{\Delta x}(x) &= \frac{\Delta f}{\Delta x}(x) + \frac{\Delta g}{\Delta x}(x) && \text{According to Sum Rule for Limits.} \\ &\to \frac{df}{dx} + \frac{dg}{dx} \quad \text{as } \Delta x \to 0\,. \end{aligned}$$

In terms of motion, if two runners are running *away* from each other starting from a common location, then the distance between them is the sum of the distances they have covered.

The formula in the Lagrange notation is as follows:

$$\boxed{(f+g)'(x) = f'(x) + g'(x)}$$

The same proof applies to *subtraction* of the change.

Exercise 4.1.4

State the Difference Rule.

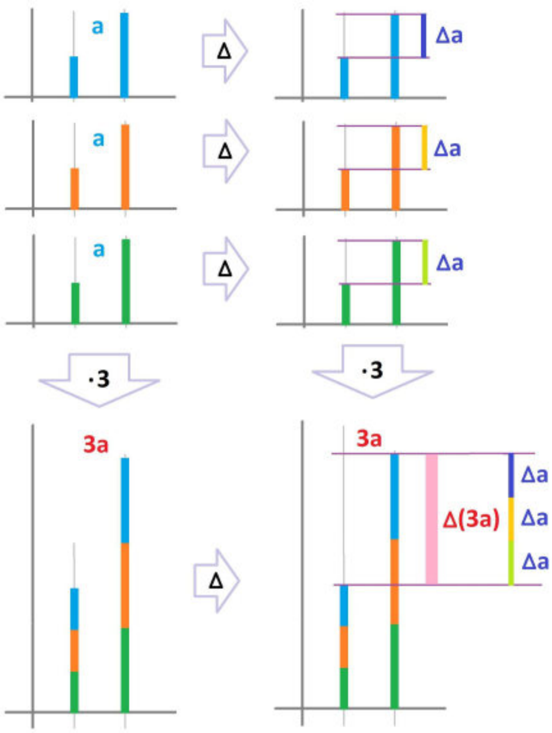
In terms of motion, if two runners are running *alongside* each other starting from a common location, then the distance between them is the difference of the distances they have covered.

Another simple statement about *motion* concerns change of units, such as from miles to kilometers:

▶ `IF` the distance is re-scaled, `THEN` so is the velocity – at the same proportion.

The velocity will have to be divided by 1000.

The idea why a *proportional* change causes the same proportional change in the differences is illustrated below (tripling):



Here, either of the two routes along the diagram produces the same result. Indeed, if the heights triple, then so do the height differences.

The algebra behind this geometry is very simple:

$$kA - ka = k(A - a)\,.$$

It's the *Distributive Rule*. This is how it applies to sequences:

Theorem 4.1.5: Constant Multiple Rule for Differences

The difference of a multiple of a function is the multiple of the function's difference.

In other words, for any function f, the its difference satisfies:

$$\boxed{\Delta(kf) = k\Delta f}$$

Proof.

Applying the definition to the function $c\,f$, we have:

$$\begin{aligned}\Delta(k\cdot f)(c) &= (k\cdot f)(x+\Delta x) - (k\cdot f)(x)\\ &= k\cdot f(x+\Delta x) - k\cdot f(x)\\ &= k\cdot\big(f(x+\Delta x) - f(x)\big)\\ &= k\cdot \Delta f\,(c)\,.\end{aligned}$$

Simply dividing this formula by Δx will produce its analog for difference quotients:

Theorem 4.1.6: Constant Multiple Rule for Difference Quotients

The difference quotient of a multiple of a function is the multiple of the function's difference quotient.

In other words, for any function f defined at the adjacent nodes x and $x+\Delta x$ of a partition and any real k, the difference quotients (defined at the corresponding

secondary node) satisfy:

$$\boxed{\frac{\Delta(kf)}{\Delta x} = k\frac{\Delta f}{\Delta x}}$$

The limit $\Delta x \to 0$ will produce the analog for derivatives:

Theorem 4.1.7: Constant Multiple Rule for Derivatives

A multiple of a function differentiable at a point is differentiable at that point, and its derivative is equal to the multiple of the function's derivative.

In other words, for any function f differentiable at x and any real k, we have at x:

$$\boxed{\frac{d(kf)}{dx} = k\frac{df}{dx}}$$

Proof.

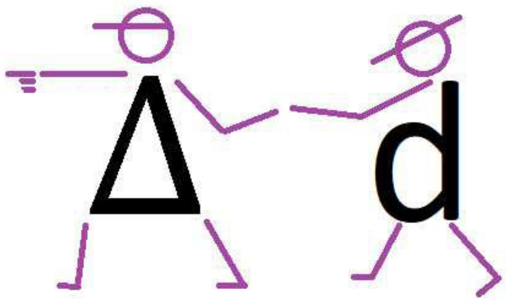
The limit with $c = x$:

$$\begin{aligned}\frac{\Delta(kf)}{\Delta x}(x) &= \frac{k\Delta f}{\Delta x}(x) \\ &= k\frac{\Delta f}{\Delta x}(x) && \text{According to Constant Multiple Rule for Limits.} \\ &\to k\frac{df}{dx}(x) \quad \text{as } \Delta x \to 0\,.\end{aligned}$$

The theorem can also be interpreted as follows:

- ▶ If the distances are proportionally increased, then so are the velocities needed to cover them, in the same period of time.

As we see (and will see again in this chapter), the derivative follows the difference and the difference quotient, every time:



The formula in the Lagrange notation is as follows:

$$\boxed{(k \cdot f)'(x) = k \cdot f'(x)}$$

Here is another way to write these formulas in the *Leibniz notation* as we omit the names of the functions. Below is the Sum Rule for functions $x \to u$, $x \to v$:

$$\boxed{\frac{d}{dx}(u+v) = \frac{du}{dx} + \frac{dv}{dx}}$$

And the Constant Multiple Rule for a function $x \to u$:

$$\boxed{\frac{d}{dx}(cu) = c\frac{du}{dx}}$$

Example 4.1.8: polynomials

We can differentiate any polynomial easily now:

$$\begin{aligned}(x^{77}+ \quad & 5x^{18} + 6x^3 - x^2 + 88)' \\ & \overset{\text{SR}}{=\!=} (x^{77})' + (5x^{18})' + (6x^3)' - (x^2)' + (88)' \\ & \overset{\text{CMR}}{=\!=} (x^{77})' + (5x^{18})' + (6x^3)' - (x^2)' + 0 \\ & \overset{\text{PF}}{=\!=} 77x^{77-1} + 5 \cdot 18x^{13-1} + 6 \cdot 3x^{3-1} - 2x^{2-1} \\ & = 77x^{76} + 90x^{17} - 18x^2 - 2x\,.\end{aligned}$$

Try to expand $(x+h)^{77}$!

These two operations can be combined into one producing *linear combinations*:

$$\alpha x + \beta y\,,$$

where α, β are two constant numbers. Recall that a function F is linear if it "preserves" linear combinations:

$$\alpha x + \beta y \quad \to \quad \boxed{\quad F \quad} \quad \to \quad \alpha F(x) + \beta F(y)$$

With this idea, these two formulas can be combined into one: The difference, the difference quotient, and the derivative are linear functions of functions. A precise version is below:

Theorem 4.1.9: Linearity of Differentiation

The difference, the difference quotient, and the derivative of a linear combination of two functions is the linear combination of their differences, the difference quotients, and the derivatives respectively, whenever they exist.

In other words, we have:

$$\boxed{\begin{array}{lll}\Delta(\alpha f + \beta g) & = \alpha \Delta f & + \beta \Delta g \\ \dfrac{\Delta(pf + \beta g)}{\Delta x} & = \alpha \dfrac{\Delta f}{\Delta x} & + \beta \dfrac{\Delta g}{\Delta x} \\ \dfrac{d(\alpha f + \beta g)}{dx} & = \alpha \dfrac{df}{dx} & + \beta \dfrac{dg}{dx}\end{array}}$$

The last formula is illustrated with the following diagram:

$$\alpha f + \beta g \quad \to \quad \boxed{\quad \frac{d}{dx} \quad} \quad \to \quad \alpha f' + \beta g'$$

Example 4.1.10: polynomials

The hierarchy of polynomials and their derivatives was used in Chapter 3 to model free fall. Let's review:

- The derivative of a constant polynomial is zero:

$$(c)' = 0\,.$$

- The derivative of a linear polynomial is constant:

$$(mx+b)' = (mx)' + (b)' = m(x)' + 0 = m \cdot 1 = m\,.$$

- The derivative of a quadratic polynomial is linear:

$$(ax^2+bx+c)' = (ax^2)' + (bx)' + (c)' = a(x^2)' + b(x)' + 0 = a \cdot 2x + b \cdot 1 = 2ax + b\,.$$

And so on. Combined with the Power Formula, the two rules above allow us to differentiate *all* polynomials. Every time, the degree goes down by 1!

The linear combinations of the power functions give us all polynomials:

Theorem 4.1.11: Derivative of Polynomial

The derivative of a polynomial of degree $n>0$ is a polynomial of degree $n-1$, as follows ($a_n \neq 0$):

$$\boxed{\begin{array}{rclllllll} f(x) & = & a_nx^n & +a_{n-1}x^{n-1} & +\ldots & +a_2x^2 & +a_1x & +a_0 \\ f'(x) & = & na_nx^{n-1} & +(n-1)a_{n-1}x^{n-2} & +\ldots & +2a_2x & +a_1 & \end{array}}$$

Exercise 4.1.12

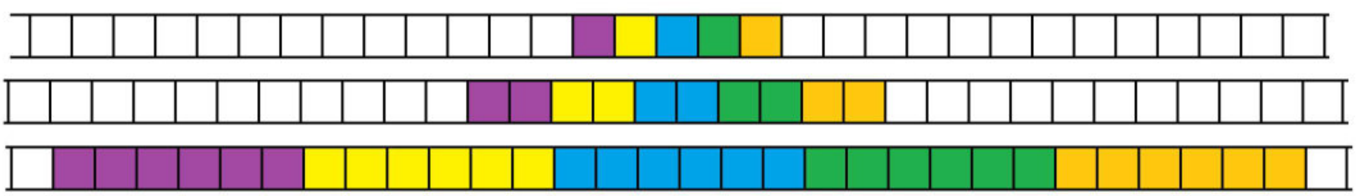
Prove the theorem.

4.2. Change of variables and the derivative

Example 4.2.1: transformations

Consider:
- The first transformation is a stretch by a factor of 2, i.e., the derivative is 2.
- The second transformation is a stretch by a factor of 3, i.e., the derivative is 3.
- Then the composition of the two transformations is a stretch by a factor of $3 \cdot 2 = 6$, i.e., the derivative is 6.

It happens as follows:



We *multiply* the derivatives.

But stretches are just *linear functions* and the stretch factors are their slopes. But the slopes are the derivatives of these functions:

	transformation	derivative
first transformation	stretch by 2	2
second transformation	stretch by 3	3
their composition	stretch by $2 \cdot 3 = 6$	$2 \cdot 3 = 6$

Exercise 4.2.2

What if the transformations are shifts? What if they are combinations of stretches and shifts?

Example 4.2.3: linear functions

Let's do the algebra. We see their derivatives and, which is the same thing for linear functions, their difference quotients:

$$\begin{array}{llllll}
 & \text{functions:} & & \text{derivatives:} & & \\
\hline
x & = qt & \Longrightarrow & \dfrac{\Delta x}{\Delta t} = \dfrac{dx}{dt} & = q & \\
 & \circ & & & \times & \\
y & = mx & \Longrightarrow & \dfrac{\Delta y}{\Delta x} = \dfrac{dy}{dx} & = m & \\
\hline
y & = m(qt) = mqt & \Longrightarrow & \dfrac{\Delta y}{\Delta t} = \dfrac{dy}{dt} & = m \cdot q & = \dfrac{\Delta x}{\Delta t} \cdot \dfrac{\Delta y}{\Delta x} = \dfrac{dx}{dt} \cdot \dfrac{dy}{dx}
\end{array}$$

In either case, we see how the intermediate variable, whether it is the difference Δx or the differential dx, is canceled or "canceled":

$$\begin{array}{ccccc}
\dfrac{\color{red}{\Delta x}}{\Delta t} & \cdot & \dfrac{\Delta y}{\color{red}{\Delta x}} & = & \dfrac{\Delta y}{\Delta t} \\
\dfrac{\color{red}{dx}}{dt} & \cdot & \dfrac{dy}{\color{red}{dx}} & = & \dfrac{dy}{dt}
\end{array}$$

Conclusion: Whichever route you take through the diagram – first composition then differentiation or the opposite – the result is the same!

Let's recall that we can interpret every composition as a change of variables. The variables of the functions we are considering are quantities we meet in everyday life. Frequently, there are multiple ways to measure these quantities:

- length and distance: inches, miles, meters, kilometers, ..., light years
- area: square inches, square miles, ..., acres
- volume: cubic inches, cubic miles, ..., liters, gallons
- time: minutes, seconds, hours, ..., years
- weight: pounds, grams, kilograms, karats
- temperature: degrees of Celsius, of Fahrenheit
- money: dollars, euros, pounds, yen
- etc.

Almost all conversion formulas are just multiplications, such as this one:

$$\#\ \text{ of meters } = \#\ \text{ of kilometers } \cdot 1000\,.$$

Warning!

We don't convert "pounds to kilos", we convert the *number of* pounds to the *number of* kilos.

Let's turn to motion:

- If the distance is measured in *inches* and time in *minutes*, the velocity is measured in *inches per minute*.
- Now, if the distance is measured in *feet*, the velocity is now measured in *feet per minute*.
- But if the time is measured in *seconds*, the velocity is measured in *inches per second*.

We are dealing with the same functions just measured in different units. How do we transition between the three?

Example 4.2.4: time units

Suppose that we have a relation (a function f) between
- x, time in minutes, and
- y, location in inches.

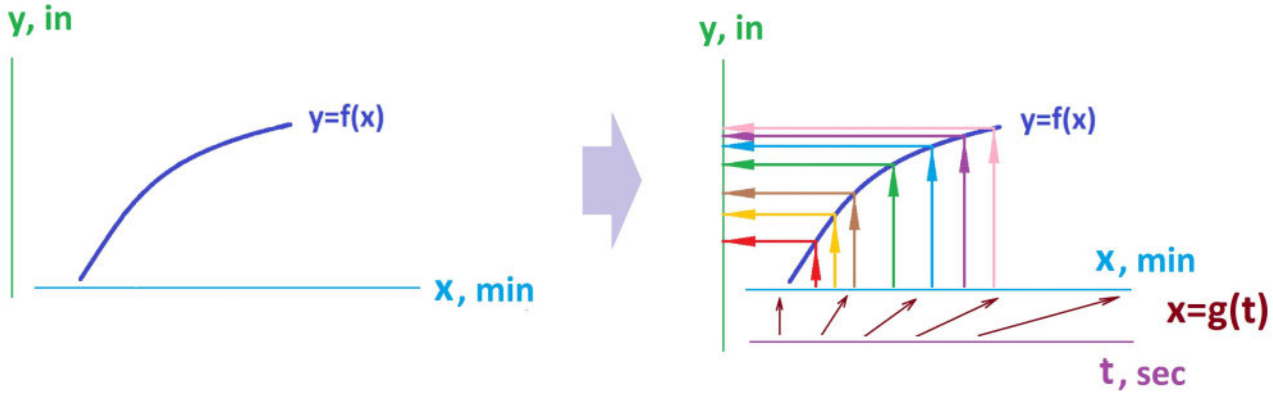
What if we need to switch to t, time in seconds? The algebra is simple:

$$x = \frac{1}{60} \cdot t \,.$$

The complete dependence is as follows:

$$t \xrightarrow{\cdot\frac{1}{60}} x \xrightarrow{f} y$$

Now, to see the new graphs, we combine the graph of f with a transformation of the x-axis, as follows:



We know that we will be covering 60 times less for every second than for a minute. That's the relation between the velocities! We write it as follows:

$$\frac{dy}{dt} = \frac{1}{60} \cdot \frac{dy}{dx} \,.$$

Example 4.2.5: distance units

Suppose again we have a relation (a function f) between
- x, time in minutes, and
- y, location in inches.

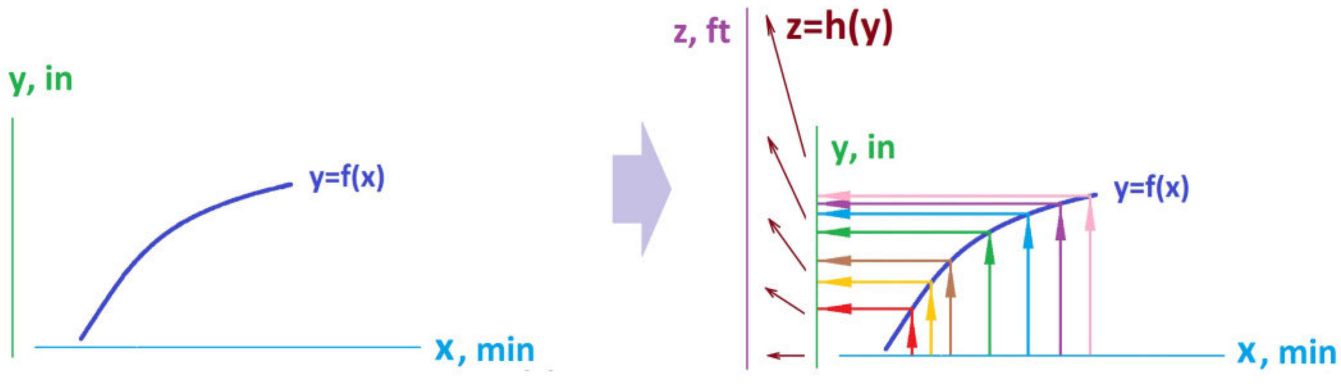
What if we need to switch to z, location in feet? The algebra is simple:

$$z = \frac{1}{12} \cdot y \,.$$

The complete dependence is as follows:

$$x \xrightarrow{f} y \xrightarrow{\cdot\frac{1}{12}} z$$

Now, to see the new graphs, we combine the graph of f with a transformation of the y-axis, as follows:



We know that we will be covering 12 inches for every foot. That's the relation between the velocities! We write it as follows:

$$\frac{dz}{dx} = \frac{1}{12} \cdot \frac{dy}{dx}.$$

Warning!

Even though we see vertical and horizontal stretching/shrinking of the original graph, it's entirely up to us to mark the units on the new axes to match the old. The graph will remain the same!

Exercise 4.2.6

What is the relation between seconds and feet?

How do such simple substitutions affect calculus as we know it?

If

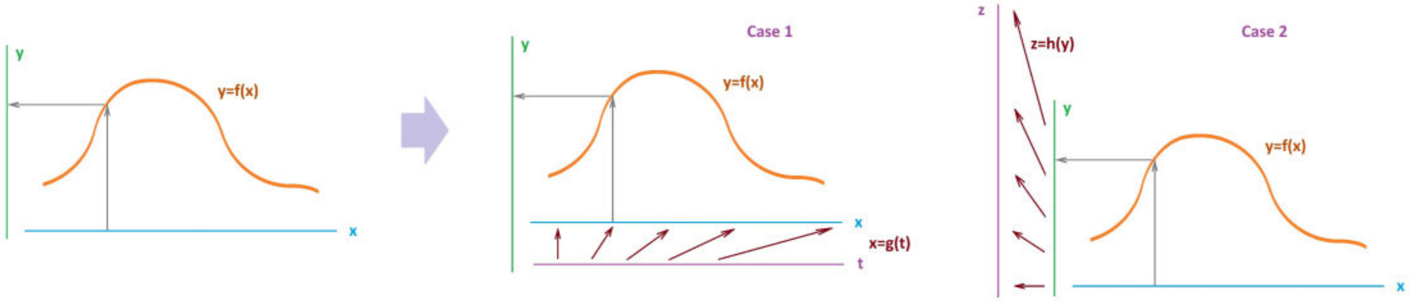
$$y = f(x)$$

is a relation between two quantities x and y, then either one may be replaced with a new variable. Let's call them t and z respectively and suppose these replacements are given by some functions:

- Case 1: $x = g(t)$
- Case 2: $z = h(y)$

These substitutions create new relations:

- Case 1: $y = k(t) = f(g(t))$
- Case 2: $z = k(x) = h(f(x))$



The graph will transform into that of the *composition* of f with this function. The *order*, however, is different. It's "before f" vs. "after f":

1. $t \to \boxed{g} \to x \to \boxed{f} \to y$
2. $x \to \boxed{f} \to y \to \boxed{h} \to z$

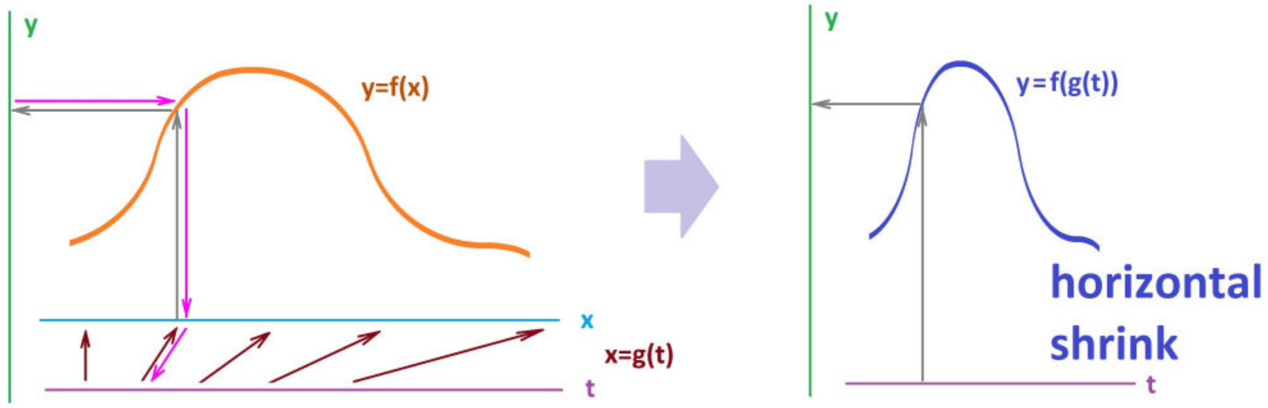
This is why the answers will be different.

This is for Case 1:

Theorem 4.2.7: Linear Chain Rule I

If $y = f(x)$ is a differentiable function and m is a real number, then we have:

$$\frac{d}{dt} f(mt) = mf'(mt)\,.$$



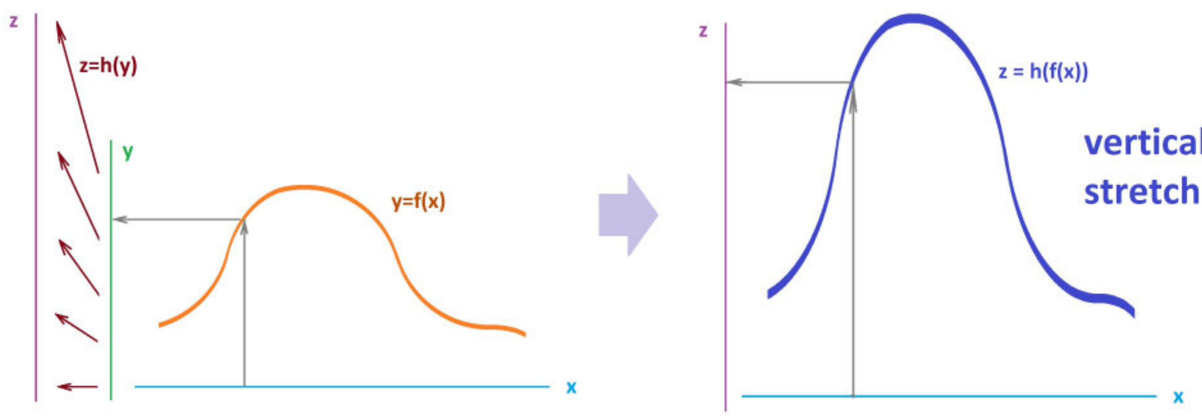
Thus, the graphs of the new quantities describing motion are simply *re-scaled* versions of the graphs of the old ones.

This is for Case 2:

Theorem 4.2.8: Linear Chain Rule II

If $y = f(x)$ is a differentiable function and m is a real number, then we have:

$$\frac{d}{dx} mf(x) = mf'(x)\,.$$



Thus, the quantities describing motion are simply replaced with their *multiples*. The new graphs are the vertically stretched versions of the old ones.

Warning!

If we are to change our unit to a logarithmic scale (for example, $x = 10^t$), then the effect on the derivative will not be proportional.

Example 4.2.9: four variables

Recall the example when we have a function f that records the temperature – in Fahrenheit – as a function f of time – in minutes – replaced with another to record the temperature in Celsius as a function g of time in seconds:

- s time in seconds
- m time in minutes
- F temperature in Fahrenheit
- C temperature in Celsius

The conversion formulas are:

$$m = s/60\,,$$

and

$$C = (F - 32)/1.8\,.$$

These are the relations between the four quantities:

$$g: \quad s \xrightarrow{\ s/60\ } m \xrightarrow{\ f\ } F \xrightarrow{\ (F-32)/1.8\ } C\,.$$

And this is the new function:

$$F = k(s) = (f(s/60) - 32)/1.8\,.$$

Then, by the *Linear Chain Rule*, we have:

$$\frac{dF}{ds} = \frac{dF}{dC}\frac{dC}{dm}\frac{dm}{ds} = \frac{1}{1.8} \cdot f'(m) \cdot \frac{1}{60}\,.$$

Exercise 4.2.10

Provide a similar analysis for the sizes of shoes and clothing.

Example 4.2.11: sine and cosine in degrees

The conversion of the number of degrees y to the number of radians x is:

$$x = \frac{\pi}{180}y\,.$$

Then,

$$\frac{dx}{dy} = \frac{\pi}{180}\,.$$

Let's denote (just once) *sine and cosine for degrees* by $\sin_d y$ and $\cos_d y$ respectively:

$$\sin_d y = \sin\left(\frac{\pi}{180}y\right) \quad \text{and} \quad \cos_d y = \cos\left(\frac{\pi}{180}y\right)\,.$$

Then,

$$\begin{aligned}\frac{d}{dy}\sin_d y &= \frac{d}{dy}\sin\left(\frac{\pi}{180}y\right)\\ &= \frac{\pi}{180}\cos\left(\frac{\pi}{180}y\right)\\ &= \frac{\pi}{180}\cos_d y\,.\end{aligned}$$

W have discovered that these function don't connect to each other as simply as the original ones: $(\sin x)' = \cos x$! Similarly, choosing a base not equal to e will ruin the intimate connection of the exponential function to its derivative: $(e^x)' = e^x$.

This is the summary of the Chain Rule:

$$\begin{tikzcd} f(g(x)) \arrow[r, "\frac{d}{dx}"] \arrow[d, "u=g(x)"] & f'(g(x))g'(x) \\ f(u) \arrow[r, "\frac{d}{du}"] & f'(u) \arrow[u, "CR"'] \end{tikzcd}$$

substitution

The method allows us to get from left to right at the top (differentiation with respect to x) by taking a detour. We follow the path around the square: substitution, differentiation with respect to u, the Chain Rule formula with back-substitution. The direct path is, of course, the limit.

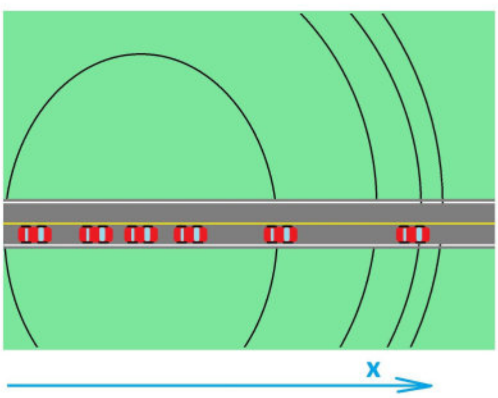
4.3. Differentiation over compositions: the Chain Rule

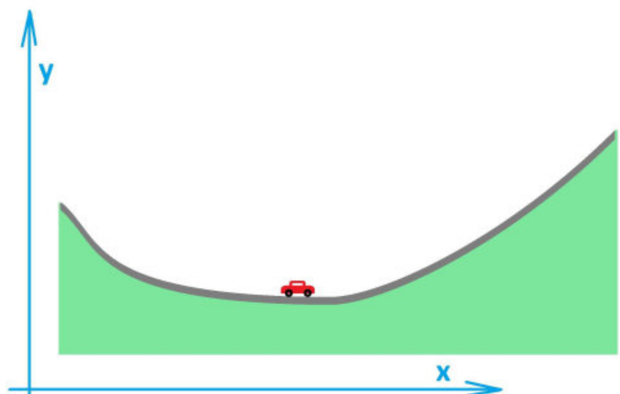
How does one express the derivative of the composition of two functions in terms of their derivatives? The answer suggested in the last section is:

▶ When you *compose* functions, you *multiply* their derivatives.

Example 4.3.1: motion

Problem: Suppose a car is driven through a mountain terrain. Its location and its speed, as seen on a map, are known. The *grade* of the road is also known. How fast is the car climbing?





We set up three variables: the time t, the (horizontal) location x, and the altitude y. There are also two known functions:

$$t \xrightarrow{f} x \xrightarrow{g} y$$

The graph of the second function g is literally the profile of the road. Their composition is what we are interested in!

We already know (from Chapter 2) that if the location, $x = f(t)$, depends continuously on time and the altitude, $z = g(x)$, continuously depends on location, then the altitude depends continuously on time as well, $z = (g \circ f)(t)$. We shall also see later that the differentiability of both functions implies the differentiability of the composition.

However, let's first dispose of the "Naive Composition Rule":

$$(f \circ g)' \overset{???}{=\!=} f' \circ g'$$

We carry out a "unit analysis" to show that such a formula simply *cannot* be true. Suppose

- t is time measured in hr.
- $x = f(t)$ is the location of the car as a function of time – measured in mi.
- $y = g(x)$ is the altitude of the road as a function of (horizontal) location – measured in ft.

- $y = h(t) = g(f(t))$ is the altitude of the road as a function of time – measured in ft.

Then,

- $f'(t)$ is the (horizontal) velocity of the car on the road – measured in $\frac{\text{mi}}{\text{hr}}$.
- $g'(x)$ is the rate of incline (slope) of the road – measured in $\frac{\text{ft}}{\text{mi}}$, with the input still measured in mi.

It doesn't even matter now what h' is measured in; just try to compose these two functions... It is impossible because the units of the output of the former and the input of the latter don't match!

However, this *is* possible:

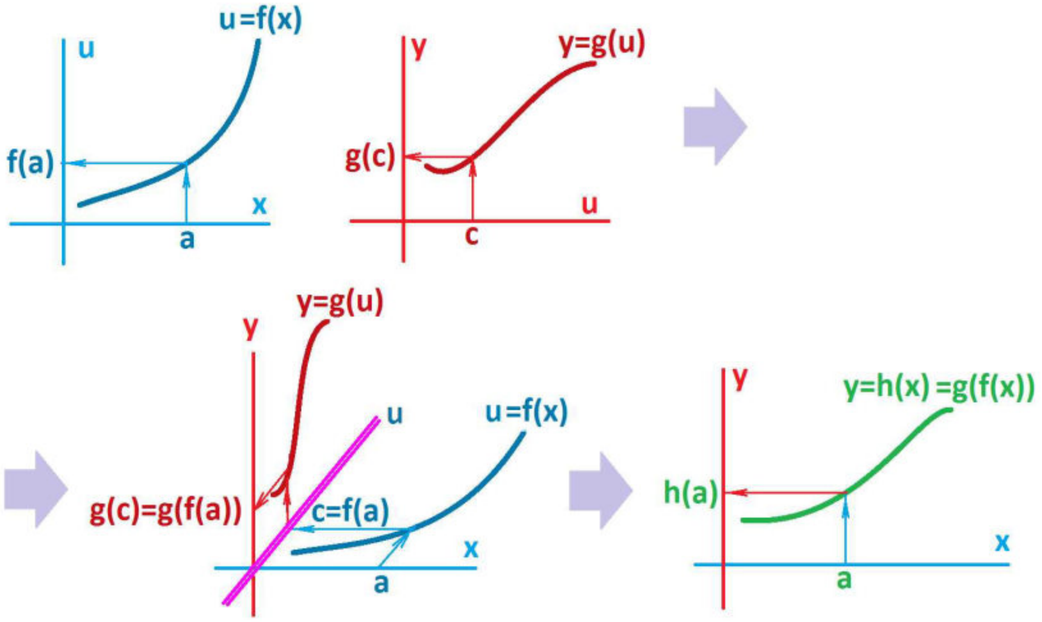
- $f'(t) \cdot g'(x)$ is their product – measured in $\frac{\text{mi}}{\text{hr}} \cdot \frac{\text{ft}}{\text{mi}} = \frac{\text{ft}}{\text{hr}}$; compared to
- $h'(t)$ is the altitude of the road as a function of time – measured in $\frac{\text{ft}}{\text{hr}}$.

This is why it makes sense:

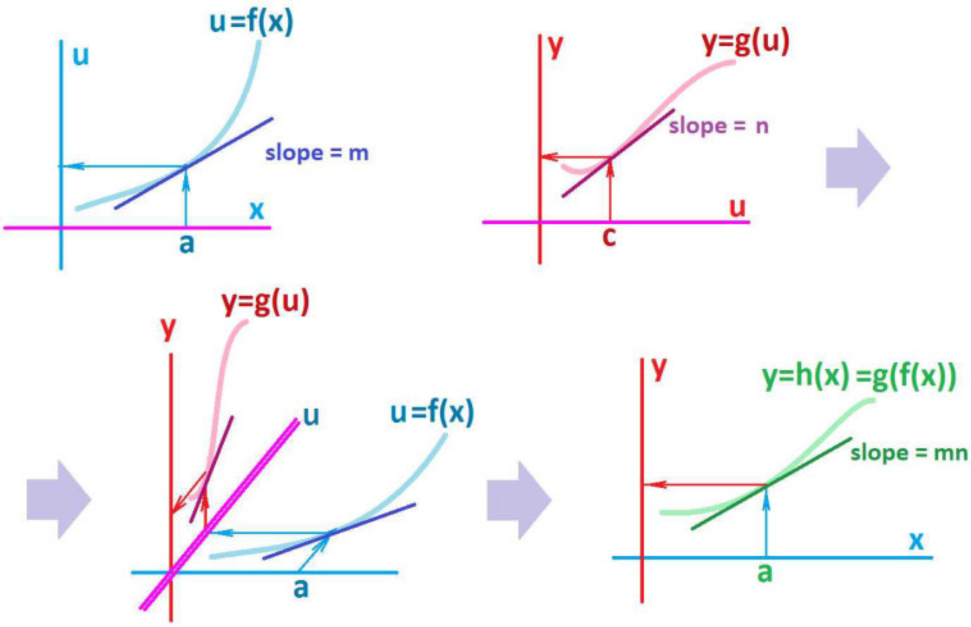
1. How fast you are climbing is proportional to your horizontal speed.
2. How fast you are climbing is proportional to the slope of the road.

This is the problem we face.

We are given two functions f and g and their composition:

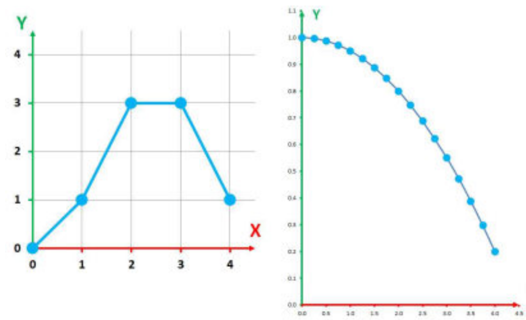


If the two derivatives are also known (possibly at a single location each), find the derivative of the composition:



According to our analysis, the slope of the composition is supposed to be the product of the slopes of the two functions! One wouldn't guess this from the picture...

We do know from the last section that the derivative of the composition of two *linear* functions is the product of the two derivatives (slopes). However, as far as single location is concerned, *all functions are linear* (approximately):



We have, therefore, strong evidence in support of our conjecture.

The dependence of the variables is as follows:

$$t \xrightarrow{\quad f \quad} x \xrightarrow{\quad g \quad} y$$

Recall that we have noticed this pattern of cancellation:

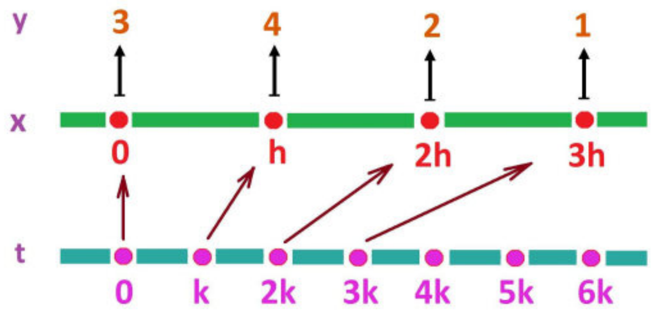
$$\begin{array}{ccccc} \frac{\color{red}{\Delta x}}{\Delta t} & \cdot & \frac{\Delta y}{\color{red}{\Delta x}} & = & \frac{\Delta y}{\Delta t} \\ \frac{\color{red}{dx}}{dt} & \cdot & \frac{dy}{\color{red}{dx}} & = & \frac{dy}{dt} \end{array}$$

Unfortunately, derivatives aren't fractions to be canceled! But difference quotients are, and we can cancel (when $\Delta x \neq 0$). We start with them.

The main difference from the rules for difference quotients we considered in the last section is that:

- There are *two* partitions: for t and for x.
- A function f must map the partition for t to the partition of x.

Below, $k = \Delta t$ and $h = \Delta x$:



We can see that, whether y depends on x or t, the differences are the same. Let's carefully state this trivial fact:

Theorem 4.3.2: Chain Rule for Differences

The difference of the composition of two functions is the difference of the second function.

In other words, for any function $x = f(t)$ defined at two adjacent nodes t and $t + \Delta t$ of a partition, and any function $y = g(x)$ defined at the two adjacent nodes $x = f(t)$ and $x + \Delta x = f(t + \Delta t)$ of a partition, we have the difference quotients (defined at the secondary nodes c and $q = f(c)$ within these edges of the two partitions respectively) satisfy:

$$\boxed{\Delta(g \circ f)(c) = \Delta g(q)}$$

To get to the difference quotients, we just multiply on the left and on the right respectively by

$$\frac{1}{\Delta t} = \frac{1}{\Delta x} \cdot \frac{\Delta x}{\Delta t},$$

producing the following:

Theorem 4.3.3: Chain Rule for Difference Quotients

The difference quotient of the composition of two functions is equal to the product of the two difference quotients.

In other words, for any function $x = f(t)$ defined at two adjacent nodes t and $t + \Delta t$ of a partition, and any function $y = g(x)$ defined at the two adjacent nodes $x = f(t)$ and $x + \Delta x = f(t + \Delta t)$ of a partition, we have the difference quotients (defined at the secondary nodes c and $q = f(c)$ within these edges of the two partitions respectively) satisfy, provided $\Delta x \neq 0$:

$$\boxed{\frac{\Delta(g \circ f)}{\Delta t}(c) = \frac{\Delta g}{\Delta x}(q) \cdot \frac{\Delta f}{\Delta t}(c)}$$

Proof.

The formula for difference quotients is deduced as follows:

$$\begin{aligned}\frac{\Delta(g\circ f)}{\Delta t}(c) &= \frac{(g\circ f)(t+\Delta t)-(g\circ f)(t)}{\Delta t}\\ &= \frac{g(f(t+\Delta t))-g(f(t))}{f(t+\Delta t)-f(t)}\,\frac{f(t+\Delta t)-f(t)}{\Delta t}\\ &= \frac{g(x+\Delta x)-g(x)}{\Delta x}\,\frac{f(t+\Delta t)-f(t)}{\Delta t}\\ &= \frac{\Delta g}{\Delta x}(q)\cdot\frac{\Delta f}{\Delta t}(c)\,.\end{aligned}$$

The limit gives us the derivative:

Theorem 4.3.4: Chain Rule for Derivatives

The composition of a function differentiable at a point and a function differentiable at the value of that point under the first function is differentiable at that point, and its derivative is equal to the product of the two derivatives.

In other words, if $x = f(t)$ is differentiable at $t = c$ and $y = g(x)$ is differentiable at $x = q = f(c)$, then we have:

$$\boxed{\frac{d(g\circ f)}{dt}(c)=\frac{dg}{dx}(q)\cdot\frac{df}{dt}(c)}$$

Proof.

Now we are to take the limit of the formula, with $c = t$, as

$$\Delta t\to 0\,.$$

Since $x = x(t)$ is continuous, we conclude that we also have: $\Delta x \to 0$. Therefore, we have:

$$\begin{array}{ccccc}\dfrac{\Delta g}{\Delta t} & = & \dfrac{\Delta g}{\Delta x}(f(t)) & \cdot & \dfrac{\Delta f}{\Delta t}(t)\\ \downarrow & & \downarrow & & \downarrow\\ \dfrac{dg}{dt} & = & \dfrac{dg}{dx}(f(t)) & \cdot & \dfrac{df}{dt}(t)\end{array}$$

The problem with the proof is: We assumed that $\Delta x \neq 0$! What if $x = f(t)$ is constant in the vicinity of t? A complete proof will be provided later.

Exercise 4.3.5

Find another, non-constant, example of a function $x = f(t)$ such that Δf may be zero even for small values of Δt.

Warning!

The input variables in the formula don't match.

The formula in the Lagrange notation is as follows:

$$\boxed{(g \circ f)'(t) = g'(f(t)) \cdot f'(t)}$$

Example 4.3.6: linear and quadratic

Find the derivative of:

$$y = (1+2x)^2 \,.$$

The function is computed in two consecutive steps (that's how we know this is a composition!):

- Step 1: From x we compute $1+2x$.
- Step 2: We square the outcome of the first step.

We then introduce an additional, disposable, variable in order to store the outcome of step 1:

$$u = 1+2x \,.$$

Then step 2 becomes:

$$y = u^2 \,.$$

This is the dependence of the variables:

$$x \to u \to y$$

Now the derivatives:

$$\begin{array}{lllll} u & = 1+2x & \Longrightarrow & \dfrac{du}{dx} & = 2 \\ y & = u^2 & \Longrightarrow & \dfrac{dy}{du} & = 2u \\ & \text{Chain Rule} & \Longrightarrow & \dfrac{dy}{dx} & = \dfrac{dy}{du} \cdot \dfrac{du}{dx} = 2u \cdot 2 = 4u \end{array}$$

Done? No. The answer must be in terms of x! Last step: Substitute $u = 1+2x$. Then the answer is $4(1+2x)$. To verify, expand, $1+4x+4x^2$, then use PF in combination with SR and CMR.

Exercise 4.3.7

Verify the result: Expand, $1+4x+4x^2$, then use PF in combination with SR and CMR.

Exercise 4.3.8

Find the derivative of $y = (2-x)^3$.

Example 4.3.9: linear and root

Now a very simple example that doesn't allow us to circumvent the *Chain Rule*. Let

$$y = \sqrt{3x+1} \,.$$

This is the abbreviated computation (decomposition, the derivatives, CR, back-substitution):

$$x \ \to\ u = 3x+1 \ \to\ y = \sqrt{u}$$

$$\underbrace{\underbrace{x \ \to\ u = 3x+1}_{\dfrac{du}{dx} = 3} \quad \underbrace{u \ \to\ y = \sqrt{u}}_{\dfrac{dy}{du} = \dfrac{1}{2\sqrt{u}}}}_{\dfrac{dy}{dx} = \dfrac{du}{dx} \cdot \dfrac{dy}{du} = 3 \cdot \dfrac{1}{2\sqrt{u}} = 3 \cdot \dfrac{1}{2\sqrt{3x+1}}}$$

Example 4.3.10: three functions

Find the derivative of:

$$z = e^{\sqrt{3x+1}}$$

Three functions this time:

$$x \ \to\ 3x+1 = u \ \to\ \sqrt{u} = y \ \to\ e^y = z$$

Fortunately, we already know a lot from the last example. We just *append* that solution with one extra step:

$$x \to u = 3x+1 \to y = \sqrt{u} \to z = e^y$$

$$\underbrace{\underbrace{\underbrace{x \ \to\ u = 3x+1}_{\dfrac{du}{dx} = 3} \quad \underbrace{u \ \to\ y = \sqrt{u}}_{\dfrac{dy}{du} = \dfrac{1}{2\sqrt{u}}}}_{\dfrac{dy}{dx} = \dfrac{du}{dx} \cdot \dfrac{dy}{du} = 3 \cdot \dfrac{1}{2\sqrt{u}}} \quad \underbrace{y \ \to\ z = e^y}_{\dfrac{dz}{dy} = e^y}}_{\dfrac{dz}{dx} = \left(\dfrac{du}{dx} \cdot \dfrac{dy}{du}\right) \cdot \dfrac{dz}{dy} = 3 \cdot \dfrac{1}{2\sqrt{u}} \cdot e^y = 3\dfrac{1}{2\sqrt{3x+1}} e^{\sqrt{3x+1}}}$$

We have applied the Chain Rule twice!

The *lesson* we have learned is: three functions – three derivatives – multiply them:

$$\boxed{\begin{array}{ccccc} & x & \to u & \to y & \to z \\ \dfrac{dz}{dx} & = \dfrac{du}{dx} & \cdot \dfrac{dy}{du} & \cdot \dfrac{dz}{dy} & \end{array}}$$

These "fractions" appear to cancel again:

$$\frac{du}{dx} \cdot \frac{dy}{du} \cdot \frac{dz}{dy} = \frac{\cancel{d}u}{dx} \cdot \frac{\cancel{d}y}{\cancel{d}u} \cdot \frac{dz}{\cancel{d}y} = \frac{dz}{dx}\,.$$

This is the *Generalized Chain Rule* about the derivative of the composition (a "chain"!) of n functions.

The short version of the Chain Rule says:

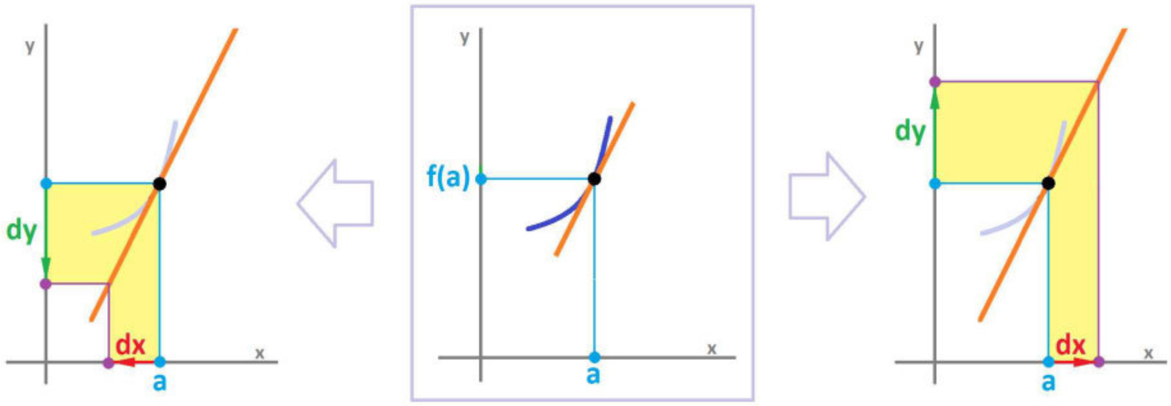
- ▶ The derivative of the composition is the product of the derivatives, as functions.

Example 4.3.11: composition of derivatives

However, if we fix the location $x = a$, we can make sense of the derivative of the composition as the *composition of the derivatives*, after all. Indeed, suppose at point a we have the derivative

$$\frac{dy}{dx} = m\,.$$

Let's, again, think of the *differentials* dx and dy as two new variables – related to each other by the above equation:



Then we think of the derivative, m, not as a number but as a *linear function*:

$$dy = m \cdot dx\,.$$

If now there is another variable t, we think of q as a linear function:

$$dx = q \cdot dt\,.$$

Then, we have to substitute q:

$$\begin{array}{lllll}
x = x(t) & = qt & \Longrightarrow & dx & = q \cdot dt \\
\quad\circ & \ \circ & & & \quad\circ \\
y = y(x) & = mx & \Longrightarrow & dy & = m \cdot dx \\
\hline
y = y(x(t)) & = m(qt) & \Longleftrightarrow & dy & = m \cdot (q \cdot dt)
\end{array}$$

We have the composition!

We can use the Chain Rule to find formulas for other important functions:

Theorem 4.3.12: Derivative of General Exponential Function

For any $b > 0$, we have:

$$\boxed{(b^x)' = b^x \ln b}$$

Proof.

We represent this exponential function in terms of the natural exponential function:

$$b^x = e^{\ln b^x} = e^{x \ln b}\,.$$

Then, by the *Chain Rule* we have:

$$(b^x)' = \left(e^{x\ln b}\right)' \overset{\text{CR}}{=\!=\!=} e^{x\ln b}\cdot (x\ln b)' = b^x \cdot \ln b\,.$$

Exercise 4.3.13

Use the idea from the proof above to find the derivative of x^x.

Example 4.3.14: trig and quadratic

Differentiate:

$$y = \sin(x^2)\,.$$

We decompose first by introducing an extra variable:

$$x \ \to\ x^2 = u \ \to\ u^2 = y\,.$$

Then, we find the derivative of the two functions and then multiply them:

$$\begin{array}{lll}
\dfrac{du}{dx} & = (x^2)' & = 2x \\
\dfrac{dy}{du} & = (\sin u)' & = \cos u \\
\hline
\dfrac{du}{dx}\cdot\dfrac{dy}{du} & & = 2x\cdot\cos u
\end{array}$$

By the Chain Rule:

$$\frac{dy}{dx} = \frac{du}{dx}\cdot\frac{dy}{du} = 2x\cdot\cos u = 2x\cos x^2\,.$$

Exercise 4.3.15

Find the derivative of $y = (\sin x)^3$. Make up your own compositions and repeat.

Let's represent the *Sum Rule*, the *Constant Multiple Rule*, and the *Chain Rule* as diagrams:

$$\begin{tikzcd}
f, g \arrow[r, "\frac{d}{dx}"] \arrow[d, "+"] \arrow[dr, phantom, "SR"] & f', g' \arrow[d, "+"] \\
f+g \arrow[r, "\frac{d}{dx}"] & (f+g)' = f' + g'
\end{tikzcd}
\qquad
\begin{tikzcd}
f \arrow[r, "\frac{d}{dx}"] \arrow[d, "\times c"] \arrow[dr, phantom, "CMR"] & f' \arrow[d, "\times c"] \\
cf \arrow[r, "\frac{d}{dx}"] & (cf)' = cf'
\end{tikzcd}
\qquad
\begin{tikzcd}
f, g \arrow[r, "\frac{d}{dx}"] \arrow[d, "\circ"] \arrow[dr, phantom, "CR"] & f', g' \arrow[d, "\times"] \\
f\circ g \arrow[r, "\frac{d}{dx}"] & (f\circ g)' = f'g'
\end{tikzcd}$$

In all of these diagrams, we start with a pair of functions at the top left and then we proceed in two ways:

- Right: Differentiate them; then down: do algebra with the results.
- Down: Do the algebra with them; then right: differentiate the result.

The result is the same!

However, we shouldn't naively assume that the same will happen to multiplication and division.

4.4. Differentiation over multiplication and division

What happens to the output function of differentiation as we perform such an algebraic operation as *multiplication* with the input functions?

Let's consider multiplication.

Example 4.4.1: tarp

Consider the following problem: If two groups of runners are unfolding a tarp (or unfurling a flag) while running east and north, respectively, what is happening to the area of this rectangle?

They may be running at different speeds:

Let's first make sure we avoid the so-called "Naive Product Rule":

$$(f \cdot g)' \overset{???}{=\!=} f' \cdot g' .$$

The formula is extrapolated from the Sum Rule but it simply *cannot* be true. Let's recast the problem in the terms of motion and take a good look at the *units*. Suppose

- x is time measured in sec.
- $y = f(x)$ is the location of the first person – measured in ft.
- $y = g(x)$ is the location of the second person – measured in ft.

Then

- $f'(x)$ is the velocity of the first person – measured in $\frac{\text{ft}}{\text{sec}}$.
- $g'(x)$ is the velocity of the second person – measured in $\frac{\text{ft}}{\text{sec}}$.

Suppose they are running in two perpendicular directions (east and north). We draw the following conclusions:

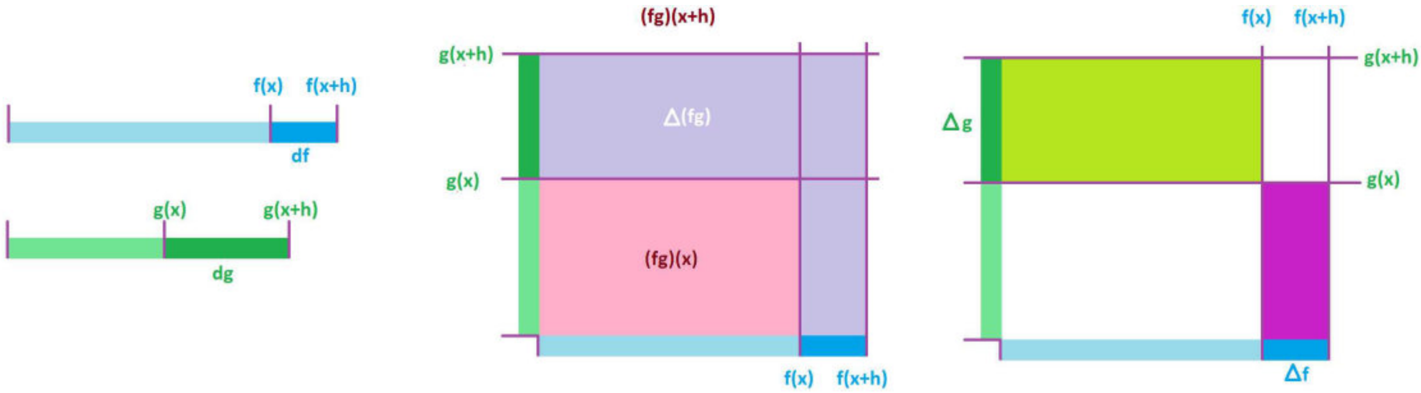
- $y = f(x) \cdot g(x)$ is the area of the rectangle enclosed by the two persons – measured in ft^2. Therefore,
- $y = (f(x) \cdot g(x))'$ is the rate of change of the area – measured in $\frac{\text{ft}^2}{\text{sec}}$. Meanwhile,
- $f(x)' \cdot g(x)'$ is an unknown quantity – measured in $\frac{\text{ft}}{\text{sec}} \cdot \frac{\text{ft}}{\text{sec}} = \frac{\text{ft}^2}{\text{sec}^2}$!

We do notice now that the product of the location and velocity gives the right units:

1. multiplied: $f'f,\ g'g$, and
2. cross-multiplied: $f'g,\ g'f$.

We conclude that a combination of some of these must make up the answer.

The product of two functions is interpreted as the *areas* of the rectangles formed by the functions, illustrated below:



As the width and the depth are increasing, so is the area of the rectangle. We can see that the increase of the area cannot be expressed entirely in terms of the increases of the width and depth! This increase is split into two parts corresponding to the two terms:

▶ When we *multiply* functions, we *cross-multiply* the functions and their derivatives.

We start with the differences:

Theorem 4.4.2: Product Rule for Differences

The difference of the product of two functions is found as the sum of the product of the function and the other function's difference.

In other words, for any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the differences (defined at the corresponding secondary node c) satisfy:

$$\boxed{\Delta(f \cdot g)(c) = f(x + \Delta x) \cdot \Delta g(c) + \Delta f(c) \cdot g(x)}$$

Proof.

The trick is to insert extra terms:

$$\begin{aligned}
\Delta(f \cdot g)(c) &= (f \cdot g)(x + \Delta x) - (f \cdot g)(x) \\
&= f(x + \Delta x) \cdot g(x + \Delta x) - f(x) \cdot g(x) \\
&= f(x + \Delta x) \cdot g(x + \Delta x) - f(x + \Delta x) \cdot g(x) + f(x + \Delta x) \cdot g(x) - f(x) \cdot g(x) \\
&= f(x + \Delta x) \cdot (g(x + \Delta x) - g(x)) + (f(x + \Delta x) - f(x)) \cdot g(x) \\
&= f(x + \Delta x) \cdot \Delta g(c) + \Delta f(c) \cdot g(x)\,.
\end{aligned}$$

Now, just divide by Δx:

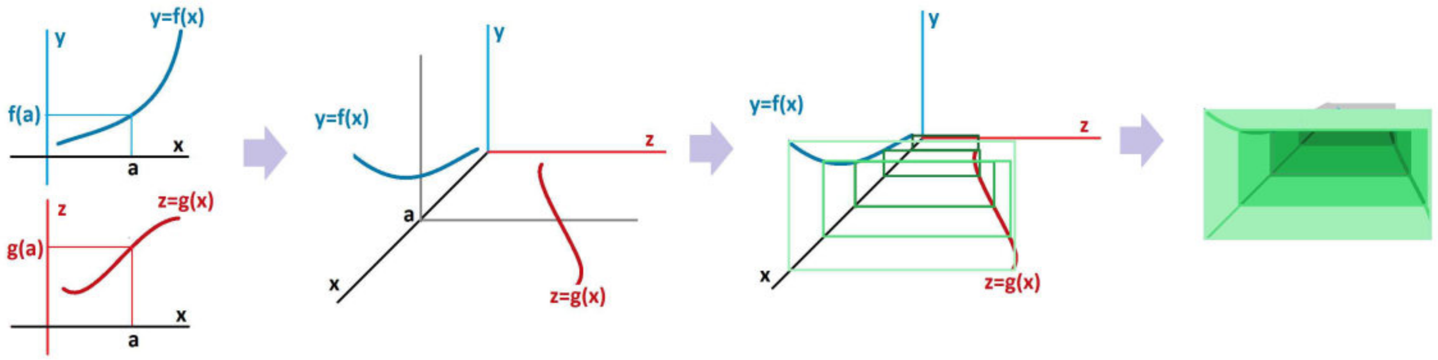
Theorem 4.4.3: Product Rule for Difference Quotients

The difference quotient of the product of two functions is found as the sum of the product of the function and the other function's difference quotient.

In other words, for any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the difference quotients (defined at the corresponding secondary node c) satisfy:

$$\boxed{\frac{\Delta(f \cdot g)}{\Delta x}(c) = f(x + \Delta x) \cdot \frac{\Delta g}{\Delta x}(c) + \frac{\Delta f}{\Delta x}(c) \cdot g(x)}$$

We already know that if the width and the height (f and g) of a rectangle are changing continuously, then so is its area ($f \cdot g$):



We shall also see that the differentiability of both dimensions implies the differentiability of the area, as follows:

Theorem 4.4.4: Product Rule for Derivatives

The product of two functions differentiable at a point is differentiable at that point and its derivative is found as the sum of the product of the function and the other function's derivative.

In other words, for any two functions f, g differentiable at x, we have:

$$\boxed{\frac{d(f\cdot g)}{dx}(x) = f(x)\cdot\frac{dg}{dx}(x) + \frac{df}{dx}(x)\cdot g(x)}$$

Proof.

The limit with $c = x$:

$$\begin{array}{cccccc} \dfrac{\Delta(f\cdot g)(x)}{\Delta x} & = f(x+\Delta x) & \cdot\dfrac{\Delta g}{\Delta x}(c) & +\dfrac{\Delta f}{\Delta x}(c) & \cdot\, g(x) & \\ & \downarrow & \downarrow & \downarrow & & \\ & f(x) & \cdot\dfrac{dg}{dx}(x) & +\dfrac{df}{dx}(x) & \cdot\, g(x) & \text{as } \Delta x\to 0\,. \end{array}$$

The first limit is justified by the fact that f, as a differentiable function, is continuous.

The formula in the Lagrange notation is as follows:

$$\boxed{(f\cdot g)'(x) = f(x)\cdot g'(x) + f'(x)\cdot g(x)}$$

In summary, this is how the cross-multiplication works:

	functions	derivatives
first	f	f'
second	g	g'

$$\longrightarrow\ (fg)' = fg' + f'g$$

Example 4.4.5: routine

Let

$$y = xe^x\,.$$

Then, we have:

$$\begin{aligned} u &= x &\implies \frac{du}{dx} &= (x)' = 1, \\ v &= e^x &\implies \frac{dv}{dx} &= (e^x)' = e^x. \end{aligned}$$

Apply *Product Rule* via "cross-multiplication", the idea of which comes from the picture above:

$$\frac{dy}{dx} = x \cdot e^x + 1 \cdot e^x = e^x(x+1)\,.$$

Next, the derivatives under *division*.

Example 4.4.6: unit analysis

Let's first make sure we avoid the so-called "Naive Quotient Rule":

$$(f/g)' \overset{???}{=\!=} f'/g'\,.$$

We can repeat the "unit analysis" to show that such a formula simply *cannot* be true. The runners still are running in two perpendicular directions, and we have:

- $y = f(x)/g(x)$ is unitless, and then
- $y = (f(x)/g(x))'$ is measured in $\frac{1}{\text{sec}}$, while
- $f(x)'/g(x)'$ is unitless!

The general case of division is more complex:

Theorem 4.4.7: Quotient Rule for Differences

For any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the differences (defined at the corresponding secondary node c) satisfy:

$$\boxed{\Delta(f/g)(c) = \frac{f(x+\Delta x)\cdot \Delta g(c) - \Delta f(c)\cdot g(x)}{g(x)g(x+\Delta x)}}$$

Proof.

We start with the case $f = 1$. Then we have:

$$\begin{aligned} \Delta(1/g)(x) &= \frac{1}{g(x+\Delta x)} - \frac{1}{g(x)} \\ &= \frac{g(x) - g(x+\Delta x)}{g(x+\Delta x)g(x)}\,. \end{aligned}$$

Now the general formula follows from the *Product Rule*.

Theorem 4.4.8: Quotient Rule for Difference Quotients

For any two functions f, g defined at the adjacent nodes x and $x + \Delta x$ of a partition, the difference quotients (defined at the corresponding secondary node c) satisfy:

$$\boxed{\frac{\Delta(f/g)}{\Delta x}(c) = \frac{f(x+\Delta x)\cdot \frac{\Delta g}{\Delta x}(c) - \frac{\Delta f}{\Delta x}(c)\cdot g(x)}{g(x)\cdot g(x+\Delta x)}}$$

provided $g(x),\ g(x+\Delta x) \neq 0$.

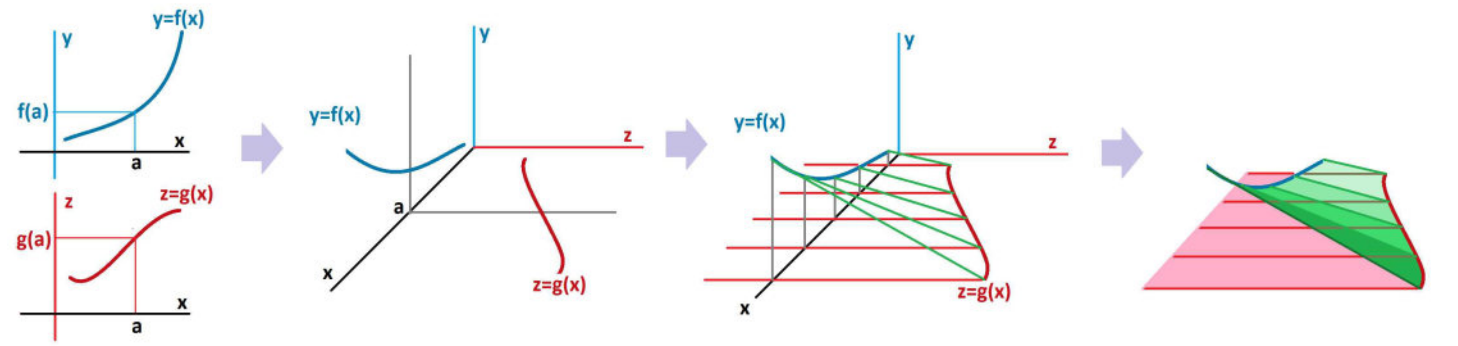
Proof.

We start with the case $f = 1$. Then we have:

$$\begin{aligned}\frac{\Delta(1/g)(x)}{\Delta x} &= \frac{g(x)-g(x+\Delta x)}{\Delta x g(x+\Delta x)g(x)}\\ &= -\frac{g(x+\Delta x)-g(x)}{\Delta x}\cdot\frac{1}{g(x+\Delta x)\cdot g(x)}\\ &= -\frac{\Delta g}{\Delta x}(c)\cdot\frac{1}{g(x+\Delta x)\cdot g(x)} &&\text{with } c = x\,.\end{aligned}$$

Now the general formula follows from the *Product Rule*.

We already know that if the width and the height (f and g) of a triangle are changing continuously, then so is the tangent of its base angle (f/g):



We shall also see that the differentiability of either dimension implies the differentiability of the tangent, as follows:

Theorem 4.4.9: Quotient Rule for Derivatives

For any two functions f, g differentiable at x, we have:

$$\boxed{\frac{d(f/g)}{dx}(x) = \frac{f(x)\cdot\frac{dg}{dx}(x) - \frac{df}{dx}(x)\cdot g(x)}{g(x)^2}}$$

provided $g(x) \neq 0$.

Proof.

We start with the case $f = 1$. Then we have:

$$\begin{aligned}\frac{\Delta(1/g)(x)}{\Delta x} &= -\frac{\Delta g}{\Delta x}(c)\cdot\frac{1}{g(x+\Delta x)\cdot g(x)} &&\text{with } c = x\\ &\to -\frac{dg}{dx}(x)\cdot\frac{1}{g(x)\cdot g(x)} &&\text{as } \Delta x \to 0\,.\end{aligned}$$

The limit of the second fraction is justified by the fact that g, as a differentiable function, is continuous. Alternatively, we represent the reciprocal of g as a composition:

$$z = \frac{1}{g(x)} \implies z = \frac{1}{y},\ y = g(x) \implies \frac{dz}{dy} = -\frac{1}{y^2},\ \frac{dy}{dx} = g'(x) \implies \frac{dz}{dx} = -\frac{1}{g(x)^2}g'(x)\,,$$

by the *Chain Rule*. Now the general formula follows from the *Product Rule*.

The formula in the Lagrange notation is as follows:

$$\boxed{\left(\frac{f(x)}{g(x)}\right)' = \frac{f'(x)\cdot g(x) - f(x)\cdot g'(x)}{g(x)^2}}$$

The formula is similar to the *Product Rule* in the sense that it also involves *cross-multiplication*:

	functions	derivatives
first	f	f'
second	g	g'

$$\longrightarrow \left(\frac{f}{g}\right)' = \frac{fg' - f'g}{g^2}$$

Example 4.4.10: derivative of tangent

$$\begin{aligned}
(\tan x)' &= \left(\frac{\sin x}{\cos x}\right)' \\
&\overset{\text{QR}}{=\!=} \frac{(\sin x)'\cos x - \sin x(\cos x)'}{(\cos x)^2} \\
&= \frac{\cos x\cos x - \sin x(-\sin x)}{\cos^2 x} \\
&= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} && \text{Use the Pythagorean Theorem.} \\
&= \frac{1}{\cos^2 x} && \text{This is } \sec^2 x\,.
\end{aligned}$$

In the Leibniz notation, this is the form of the *Product Rule*:

$$\boxed{\frac{d}{dx}(uv) = \frac{du}{dx}\cdot v + \frac{dv}{dx}\cdot u}$$

and the Quotient Rule:

$$\boxed{\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{\frac{du}{dx}\cdot v - \frac{dv}{dx}\cdot u}{v^2}}$$

Example 4.4.11: Power Formula proof

In the last chapter, we proved the Power Formula for *positive* integers:

$$(x^m)' = mx^{m-1}\,.$$

Let's prove the negative integers. We treat the function as a fraction:

$$x^{-m} = \frac{1}{x^m}\,.$$

We differentiate the numerator and the denominator:

$$\begin{array}{lll}
u = 1 & \Longrightarrow & u' = 0 \\
v = x^m & \Longrightarrow & v' = mx^{m-1} \quad \text{According to the Product Formula.}
\end{array}$$

We apply the Quotient Rule now:

$$\begin{aligned}
(x^{-m})' &= \left(\frac{1}{x^m}\right)' \\
&= \frac{0 \cdot x^m - 1 \cdot mx^{m-1}}{(x^m)^2} \\
&= -\frac{mx^{m-1}}{(x^m)^2} \\
&= -m\frac{x^{m-1}}{x^{2m}} \\
&= -mx^{-m-1}\,.
\end{aligned}$$

If we replace $-m$ with n, we have the familiar formula:

$$(x^n)' = nx^{n-1}\,.$$

Example 4.4.12: ratio

Find

$$\left(\frac{\sqrt{x}}{x^2+1}\right)'\,.$$

Consider:

$$\begin{aligned}
u &= \sqrt{x} &\implies \frac{du}{dx} &= \frac{1}{2\sqrt{x}} \\
v &= x^2+1 &\implies \frac{dv}{dx} &= 2x
\end{aligned}$$

Then,

$$\frac{d}{dx}\left(\frac{u}{v}\right) = \frac{\frac{1}{2\sqrt{x}}(x^2+1) - \sqrt{x}\cdot 2x}{(x^2+1)^2}\,.$$

No need to simplify.

Example 4.4.13: from limit to derivative

This is a different kind of example. Evaluate:

$$\lim_{x\to 5}\frac{2^x-32}{x-5}\,.$$

It's just a limit. But we recognize that this is the derivative of some function. We compare the expression to the formula in the definition:

$$f'(a) = \lim_{x\to a}\frac{f(x)-f(a)}{x-a}\,,$$

and match. So, we have here:

$$a=5,\ f(x)=2^x,\ f(5)=2^5=32\,.$$

Therefore, our limit is equal to $f'(5)$ for $f(x)=2^x$. Compute:

$$f'(x) = (2^x)' = 2^x\ln 2\,,$$

so

$$\lim_{x\to 5}\frac{2^x-32}{x-5} = f'(5) = 2^5\ln 2 = 32\ln 2\,.$$

4.5. The rate of change of the rate of change

We saw in Chapter 3 how to derive the acceleration from the location, as two difference quotients:

$$\text{location} \xrightarrow{\text{DQ}} \text{velocity} \xrightarrow{\text{DQ}} \text{acceleration}$$

Furthermore, if a function is known at the nodes of a partition, its difference quotient is also a function – known at the secondary nodes. Can we treat the latter as a function too?

Example 4.5.1: acceleration

Suppose the time is increasing by 1 so that we only need to look at the *differences*. We progress from the locations defined at the nodes of the partition of the time line to the velocities defined on the edges to the accelerations defined at the nodes again:

location:	–	2	– – –	5	– – –	10	–
velocity:	–	– • –	$5-3=3$	– • –	$10-5=5$	– • –	–
acceleration:	–	–?–	– – –	$5-3=2$	– – –	–?–	–
time:		1.0	1.5	2.0	2.5	3.0	

Generally, if we know only *three* values of a function (first line), we can compute the difference quotients along the two intervals (second line), and then place the results at the corresponding edge:

function:	–	$f(x_1)$	– – –	$f(x_2)$	– – –	$f(x_3)$	–
DQ:	–	– • –	$\dfrac{\Delta f}{\Delta x_2}$	– • –	$\dfrac{\Delta f}{\Delta x_3}$	– • –	–
DQ of DQ:	–	–?–	– – –	$\dfrac{\frac{\Delta f}{\Delta x_3}-\frac{\Delta f}{\Delta x_2}}{c_3-c_2}$	– – –	–?–	–
variable:		x_1	c_2	x_2	c_3	x_3	

To find the change of this new function, we carry out the same operation and place the result in the middle (third line).

Let's review the construction of the difference quotient.

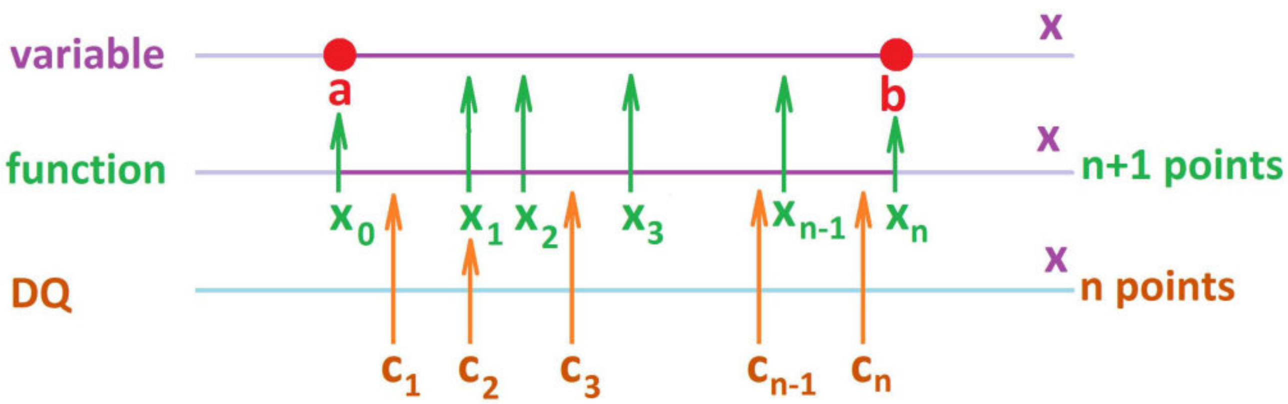
First, we have an *augmented partition* of an interval $[a,b]$. We partition it into n intervals with the help of the nodes (the end-points of the intervals):

$$a=x_0,\ x_1,\ x_2,\ ...,\ x_{n-1},\ x_n=b\,.$$

We also provide the secondary nodes:

$$c_1 \text{ in } [x_0,x_1],\ c_2 \text{ in } [x_1,x_2],\ ...,\ c_n \text{ in } [x_{n-1},x_n]\,.$$

A function is defined on the former, and its difference quotient on the latter:

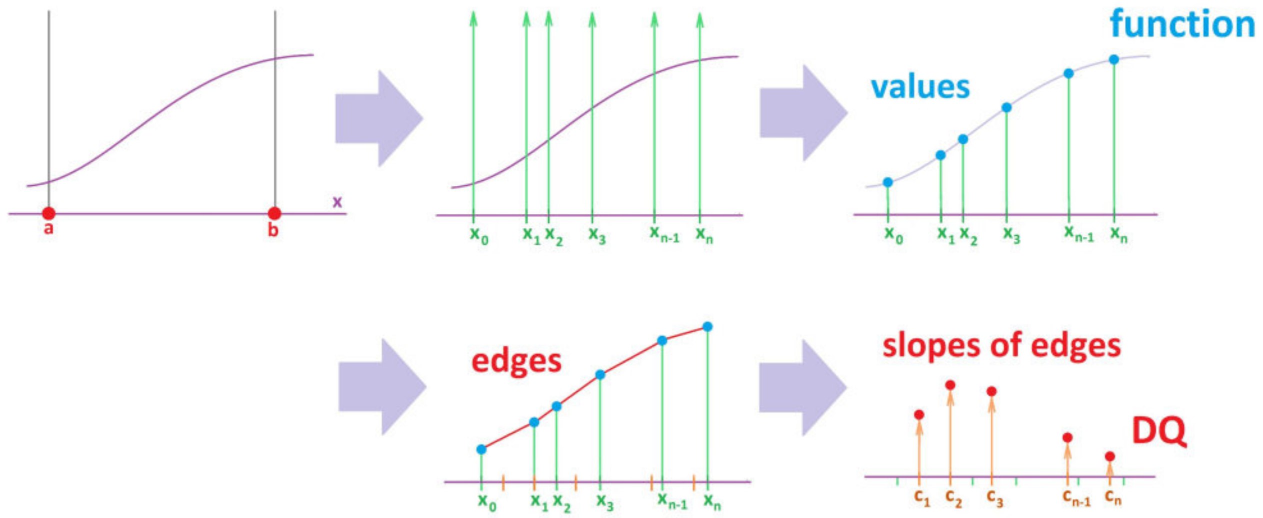


Suppose a function $y = f(x)$ is defined at the nodes $x_k,\ k = 0, 1, 2, ..., n$. Then the *difference quotient* of f is defined at the secondary nodes of the partition by the following:

$$\boxed{\frac{\Delta f}{\Delta x}(c_k) = \frac{f(x_{k+1}) - f(x_k)}{x_{k+1} - x_k}}$$

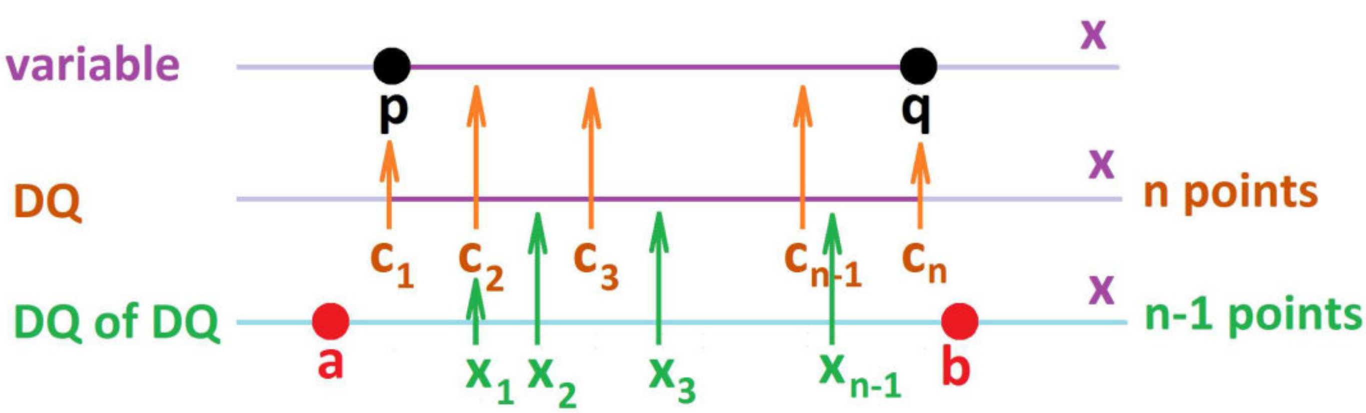
for each $k = 1, 2, ..., n$.

The new function represents the slopes of the secant lines over the nodes of the partition:



It is now especially important that we have utilized the secondary nodes as the inputs of the new function. Indeed, we can now carry out a similar construction with this new function and find the the rate of change of the rate of change.

We are in the same position as when we started: We have now a new *augmented partition*:



The complete construction is as follows. The interval is

$$[p, q], \quad \text{with } p = c_0 \ \text{ and } \ q = c_n\,.$$

We partition it into $n-1$ intervals with the help of the nodes that used to be the secondary nodes in the original partition:

$$p = c_1,\ c_2,\ c_3,\ ...,\ c_{n-1},\ c_n = b\,.$$

Then the increments are:

$$\Delta c_k = c_{k+1} - c_k\,.$$

Now, what are the secondary nodes? The primary nodes of the last partition of course! Indeed, we have:

$$x_1 \text{ in } [c_1, c_2],\ x_2 \text{ in } [c_2, c_3],\ ...,\ x_{n-1} \text{ in } [c_{n-1}, c_n]\,.$$

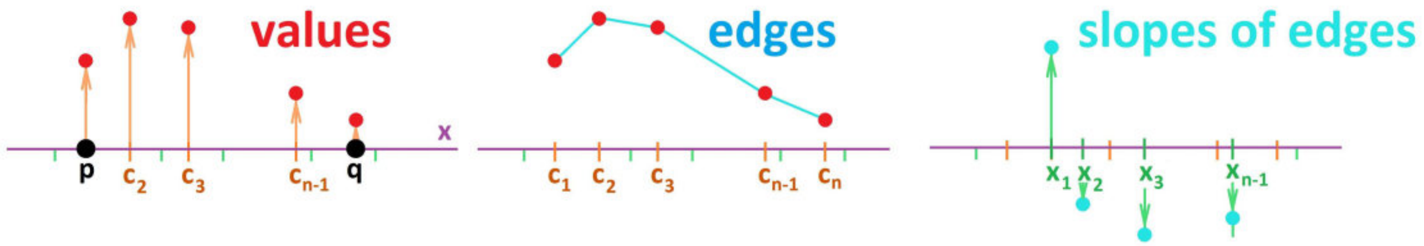
We apply the same construction to the function

$$g = \frac{\Delta f}{\Delta x}\,.$$

The difference quotient of g is defined at the secondary nodes of the new partition by the following:

$$\boxed{\frac{\Delta g}{\Delta x}(x_k) = \frac{g(c_{k+1}) - g(c_k)}{c_{k+1} - c_k}}$$

for each $k = 1, 2, ..., n-1$. It is visualized as follows:



We put the two formulas together:

Definition 4.5.2: second difference quotient

The *second difference quotient* of a function f defined at the primary nodes of a partition is defined to be the difference quotient of the difference quotient; i.e., it is defined at the primary nodes of the partition (and denoted) as follows:

$$\boxed{\frac{\Delta^2 f}{\Delta x^2}(x_k) = \frac{\frac{\Delta f}{\Delta x}(c_{k+1}) - \frac{\Delta f}{\Delta x}(c_k)}{c_{k+1} - c_k}}$$

for each $k = 1, 2, ..., n-1$.

There are:

- $n+1$ values of f (at the primary nodes),
- n values of $\dfrac{\Delta f}{\Delta x}$ (at the secondary nodes), and
- $n-1$ values of $\dfrac{\Delta^2 f}{\Delta x^2}$ (at the primary nodes except a and b).

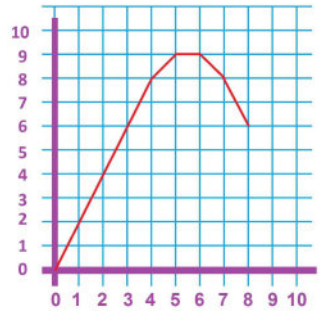
We will often omit the subscripts for the simplified notation:

Second difference quotient

$$\frac{\Delta^2 f}{\Delta x^2}(x) = \frac{\frac{\Delta f}{\Delta x}(c + \Delta c) - \frac{\Delta f}{\Delta x}(c)}{\Delta c}$$

Example 4.5.3: curvature

As we know, the difference quotient of a linear function is constant. The second difference quotient is, therefore, zero. We conclude that a non-zero second difference quotient indicates a non-linear graph:



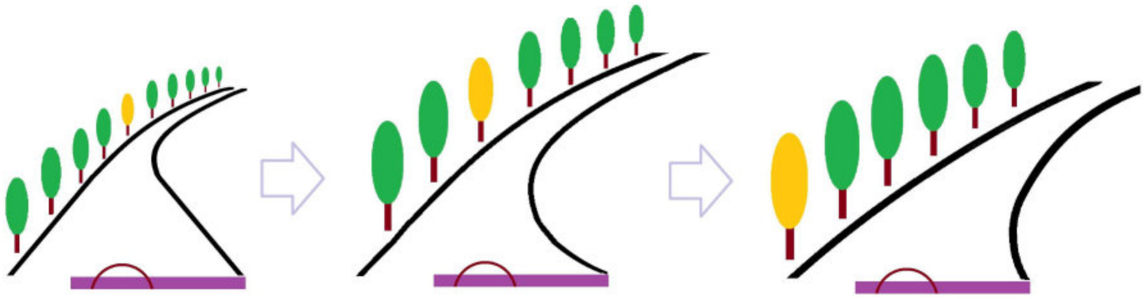
Above, the slopes remain the same, 2, at first; therefore, the second difference quotient is zero:

$$\frac{\Delta^2 f}{\Delta x^2} = 0\,.$$

Then, the slope changes to 1, and this change is the second difference quotient (assuming $\Delta x = 1$):

$$\frac{\Delta^2 f}{\Delta x^2} = -1\,.$$

As another way to see this idea, imagine yourself driving along a straight part of the road and seeing a particular tree ahead (no curvature), then, as you start to turn, the trees start to pass your field of vision from right to left (curvature):



Furthermore, higher values of the second difference quotient mean higher *curvature* of the graph of $y = f(x)$.

Example 4.5.4: sine and cosine

Let's differentiate $\sin x$ for the second time. In Chapter 3, we found its difference quotient over a mid-point partition with a single interval. This time we will need at least two intervals:

- three nodes x: $a-h$, a, and $a+h$, and
- two secondary nodes c: $a-h/2$ and $a+h/2$.

We use the two formulas for the difference quotients of $\sin x$ and $\cos x$ from Chapter 3. We write the former for the two secondary nodes, but we rewrite the latter for the partition with two nodes $a-h/2$, $a+h/2$ and a single secondary node $x=a$:

$$\frac{\Delta}{\Delta x}(\sin x) \quad = \frac{\sin(h/2)}{h/2}\cdot \cos c\,, \quad \frac{\Delta}{\Delta x}(\cos x) = -\frac{\sin(h/2)}{h/2}\cdot \sin a\,.$$

Therefore, we have at a:

$$\begin{aligned}
\frac{\Delta^2}{\Delta x^2}(\sin x) &= \frac{\Delta}{\Delta x}\left(\frac{\Delta}{\Delta x}(\sin x)\right)(a) \\
&= \frac{\Delta}{\Delta x}\left(\frac{\sin(h/2)}{h/2}\cdot\cos c\right) && \text{According to the first formula.} \\
&= \frac{\sin(h/2)}{h/2}\frac{\Delta\cos}{\Delta x}(a) && \text{According to CMR.} \\
&= \frac{\sin(h/2)}{h/2}\left(-\frac{\sin(h/2)}{h/2}\cdot\sin a\right) && \text{According to the second formula.} \\
&= -\left(\frac{\sin(h/2)}{h/2}\right)^2\cdot\sin a\,.
\end{aligned}$$

It's just the original function upside down and stretched! Similarly, we find:

$$\frac{\Delta}{\Delta x}(\cos x) = -\frac{\sin(h/2)}{h/2}\cdot\sin c \implies \frac{\Delta^2}{\Delta x^2}(\cos x) = -\left(\frac{\sin(h/2)}{h/2}\right)^2\cdot\cos a\,.$$

Example 4.5.5: exponential function

For the exponential function, we need a left-end partition with two intervals:

- three nodes x: $a-h$, a, and $a+h$, and
- two secondary nodes c: $a-h$ and a.

Then, we find at a:

$$\frac{\Delta}{\Delta x}(e^x) = \frac{e^h-1}{h}e^{-h/2}\cdot e^c \implies \frac{\Delta^2}{\Delta x^2}(e^x) = \left(\frac{e^h-1}{h}e^{-h/2}\right)^2\cdot e^a\,.$$

It's just the original function stretched vertically!

This construction will be repeatedly used for approximations and simulations. It will be followed, when necessary, by taking its limit:

Definition 4.5.6: second derivative

The *second derivative* of a function f is defined to be limit of the second difference quotient:

$$\boxed{\frac{d^2f}{dx^2}(x) = \lim_{\Delta x\to 0}\frac{\Delta^2 f}{\Delta x^2}(x_k) = \lim_{\Delta x\to 0}\frac{\frac{\Delta f}{\Delta x}(c_{k+1}) - \frac{\Delta f}{\Delta x}(c_k)}{c_{k+1}-c_k}}$$

We accept the following without proof:

Theorem 4.5.7: Second Derivative

The second derivative is the derivative of the derivative:

$$\boxed{\frac{d^2f}{dx^2} = (f')'}$$

As we repeatedly differentiate the same function f, we just add more and more "primes":

$$f'',\ f''',\ f'''',\ \dots$$

4.6. Repeated differentiation

Example 4.6.1: $\sin$

Let's continue to differentiate the sine:

$$\begin{array}{llll}
(\sin x)' & = \cos x & & \\
(\cos x)' & = -\sin x & \Longrightarrow & (\sin x)'' = -\sin x \\
(-\sin x)' & = -\cos x & \Longrightarrow & (\sin x)''' = -\cos x \\
(-\cos x)' & = \sin x & \Longrightarrow & (\sin x)'''' = \sin x
\end{array}$$

And we are back where we started, i.e., the differentiation process for this particular function is cyclic!

We use the following terminology and notation:

Consecutive derivatives

function	f	$f^{(0)}$	
first derivative	f'	$f^{(1)}$	$\dfrac{df}{dx}$
second derivative	$f''=(f')'$	$f^{(2)}=\left(f^{(1)}\right)'$	$\dfrac{d^2f}{dx^2}=\dfrac{d}{dx}\left(\dfrac{df}{dx}\right)$
third derivative	$f'''=(f'')'$	$f^{(3)}=\left(f^{(2)}\right)'$	$\dfrac{d^3f}{dx^3}=\dfrac{d}{dx}\left(\dfrac{d^2f}{dx^2}\right)$
...		...	...
nth derivative		$f^{(n)}=\left(f^{(n-1)}\right)'$	$\dfrac{d^nf}{dx^n}=\dfrac{d}{dx}\left(\dfrac{d^{n-1}f}{dx^{n-1}}\right)$
...		...	...

Thus, a given differentiable function may produce a *sequence of functions*:

$$f \to \boxed{\frac{d}{dx}} \to f' \to \boxed{\frac{d}{dx}} \to f'' \to \ldots \to f^{(n)} \to \ldots$$

or

$$f \xrightarrow{\frac{d}{dx}} f' \xrightarrow{\frac{d}{dx}} f'' \xrightarrow{\frac{d}{dx}} \ldots \xrightarrow{\frac{d}{dx}} f^{(n)} \xrightarrow{\frac{d}{dx}} \ldots$$

We just need the outcome of each step to be differentiable as well!

Note that, for a fixed x, the sequence of numbers:

$$f(x),\ f'(x),\ f''(x),\ \ldots,\ f^{(n)}(x),\ \ldots$$

is just that, a *sequence*, a concept familiar from Chapter 1. However, as the example of $\sin x$ shows, this sequence doesn't have to converge:

$$(\sin x)^{(n)}\Big|_{x=0},\ n=0,1,2,3,\ldots \text{ are equal to } 0,-1,0,1,0,\ldots$$

(We will see in Volume 3, Chapter 3IC-5, that some special linear combinations of the derivatives do produce a sequence convergent to the function.)

Let's try to compute as many consecutive derivatives as possible, or even all of them, for the functions below.

Example 4.6.2: positive integer powers

The positive integer powers are handled with the Product Formula:

$$(x^n)' = nx^{n-1}\,.$$

The power decreases by 1 every time. Therefore,

$$(x^n)^{(n+1)} = 0\,.$$

Then, it stays 0:

$$(x^n)^{(n+1)} = (x^n)^{(n+2)} = ... = 0\,.$$

The powers in the sequence of consecutive derivatives decrease to 0 and then remain constant.

Example 4.6.3: exponent

The exponential function next. Since

$$(e^x)' = e^x\,,$$

we have:

$$(e^x)^{(n)} = e^x\,.$$

The function remains the same! The sequence of consecutive derivatives is constant.

Example 4.6.4: sine and cosine

The trig functions. Same for both sine and cosine:

$$\begin{array}{ll}(\sin x)^{(4n)} & = \sin x\\ (\cos x)^{(4n)} & = \cos x\end{array}$$

The sequence of consecutive derivatives is cyclic for both functions.

Example 4.6.5: negative integer powers

The negative integer powers are subjected to the Power Formula again:

$$\begin{array}{ll}(x^{-1})' & = -1x^{-2}\\ (-x^{-2})' & = 2x^{-3}\\ ... & \end{array}$$

The power goes down by 1 every time and, as a result, tends to $-\infty$. The sequence of consecutive derivatives doesn't stop.

Exercise 4.6.6

Show that the same happens with all non-integer powers.

Differentiation produces a different dynamics for different functions:

Repeated Differentiation of Basic Functions

$$\begin{tikzcd} x^n \arrow[r, "\frac{d}{dx}"] & nx^{n-1} \arrow[r, "\frac{d}{dx}"] & \dots \arrow[r, "\frac{d}{dx}"] & \text{constant} \arrow[r, "\frac{d}{dx}"] & 0 \arrow[loop above, "\frac{d}{dx}"] \end{tikzcd}$$

$$\begin{tikzcd} \frac{1}{n} \arrow[r, "\frac{d}{dx}"] & -\frac{1}{n^2} \arrow[r, "\frac{d}{dx}"] & \frac{2}{n^3} \arrow[r, "\frac{d}{dx}"] & \dots & \text{divergent} \end{tikzcd}$$

$$\begin{tikzcd} \sin x \arrow[r, "\frac{d}{dx}"] & \cos x \arrow[d, "\frac{d}{dx}"] & \text{periodic} \\ -\cos x \arrow[u, "\frac{d}{dx}"] & -\sin x \arrow[l, "\frac{d}{dx}"'] & \end{tikzcd}$$

$$\begin{tikzcd} e^{-x} \arrow[r, leftrightarrow, "\frac{d}{dx}"] & -e^{-x} & \text{alternating} \end{tikzcd}$$

$$\begin{tikzcd} e^{x} \arrow[loop above, "\frac{d}{dx}"] & \text{constant} \end{tikzcd}$$

Warning!

Starting in Volume 4, Chapter 4HD-2, we will see that the function and its derivative are two animals of very different breeds. As a result, the dynamics discussed above will disappear in higher dimensions.

The repeated differentiation process may fail to continue when the nth derivative is *not* differentiable, i.e., when the following limit does not exist:

$$f^{(n)}(a) = \lim_{h \to 0} \frac{f^{(n-1)}(a+h) - f^{(n-1)}(a)}{h}\,.$$

Below is the summary of this idea:

Definition 4.6.7: n times differentiable function

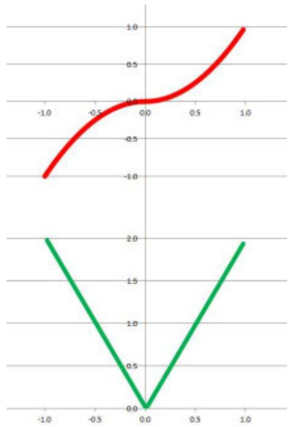
A function f is called *twice, thrice, ..., n times differentiable* when $f', f'', f''', ..., f^{(n)}$ exists respectively. When these derivatives are also continuous, we call the function n *times continuously differentiable*.

Example 4.6.8: not twice differentiable

This function is differentiable:

$$f(x) = \begin{cases} -x^2 & \text{if } x < 0\,, \\ x^2 & \text{if } x \geq 0\,. \end{cases}$$

Its graph looks smooth and there is no doubt in which direction a beam of light would bounce off such a surface:



But the graph of its derivative doesn't look smooth!

Let's compute the derivatives. It is easy for $x \neq 0$ because there is only one formula:

$$f(x) = \begin{cases} -2x & \text{if } x < 0\,, \\ 2x & \text{if } x > 0\,. \end{cases}$$

For the case of $x = 0$, we consider the two one-sided limits:

$$\lim_{h\to 0^-} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0^-} \frac{f(h)}{h} = \lim_{h\to 0} \frac{-h^2}{h} = \lim_{h\to 0}(-h) = 0$$

$$\lim_{h\to 0^+} \frac{f(0+h) - f(0)}{h} = \lim_{h\to 0^+} \frac{f(h)}{h} = \lim_{h\to 0} \frac{h^2}{h} = \lim_{h\to 0} h = 0$$

They match! Therefore,

$$f'(0) = 0\,.$$

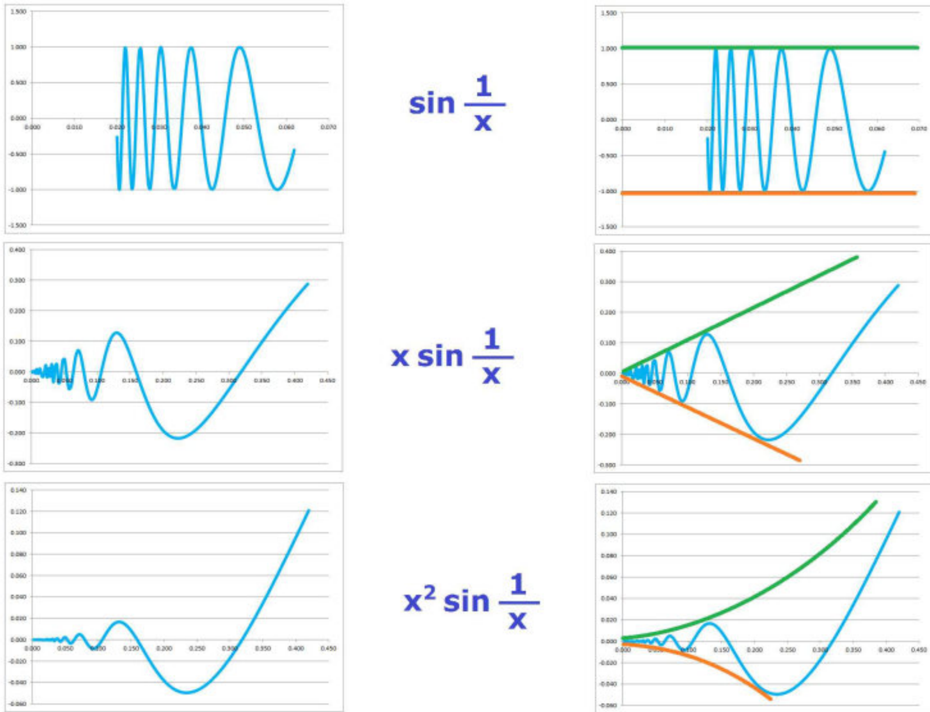
We have discovered that $f'(x) = 2|x|$. It's not differentiable at 0! The original function isn't twice differentiable.

Example 4.6.9: hierarchy

More examples of this kind:

1. $f(x) = \sin\dfrac{1}{x}$ is discontinuous at $x = 0$.
2. $g(x) = x\sin\dfrac{1}{x}$ is continuous at $x = 0$ but not differentiable.
3. $h(x) = x^2\sin\dfrac{1}{x}$ is differentiable at $x = 0$ but not twice differentiable.

We are assuming also that the value at 0 is 0 for each of them. We plot them below:



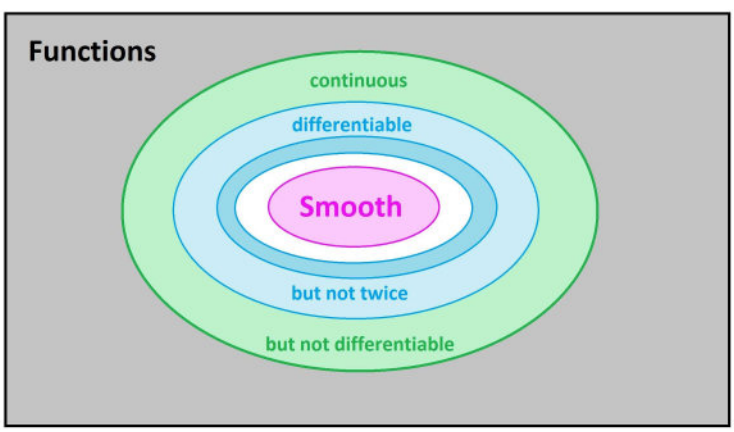
This is the reasoning:

1. The two lines don't form a squeeze.
2. The two lines do form a squeeze and it guarantees continuity at $x = 0$. But the two lines are also two secant lines with different slopes; therefore, we have no differentiability at $x = 0$.
3. The two curves form a squeeze and guarantee continuity at $x = 0$. But they also have the same slope, 0, at $x = 0$! This guarantees that the slopes of the secant lines will converge to 0; therefore, we have differentiability at $x = 0$. It might not be twice differentiable though.

Exercise 4.6.10

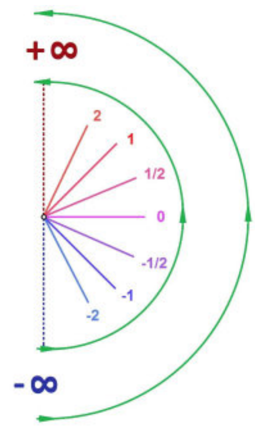
Prove the above statements.

Below we visualize the relation between these classes of functions:



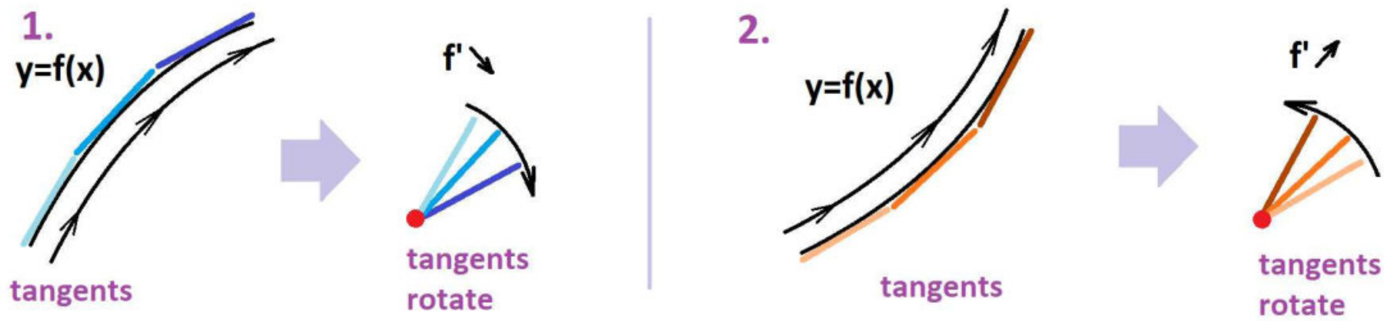
What is the geometric meaning of these *higher derivatives* for a given function?

Let's consider the first derivative. It represents the slopes of the function. Then the second derivative represents the rate of change of these slopes. Consider the wheel of slopes:



We can imagine the tangent lines rotate.

This is how changing slopes are seen as rotating tangents in the right direction:

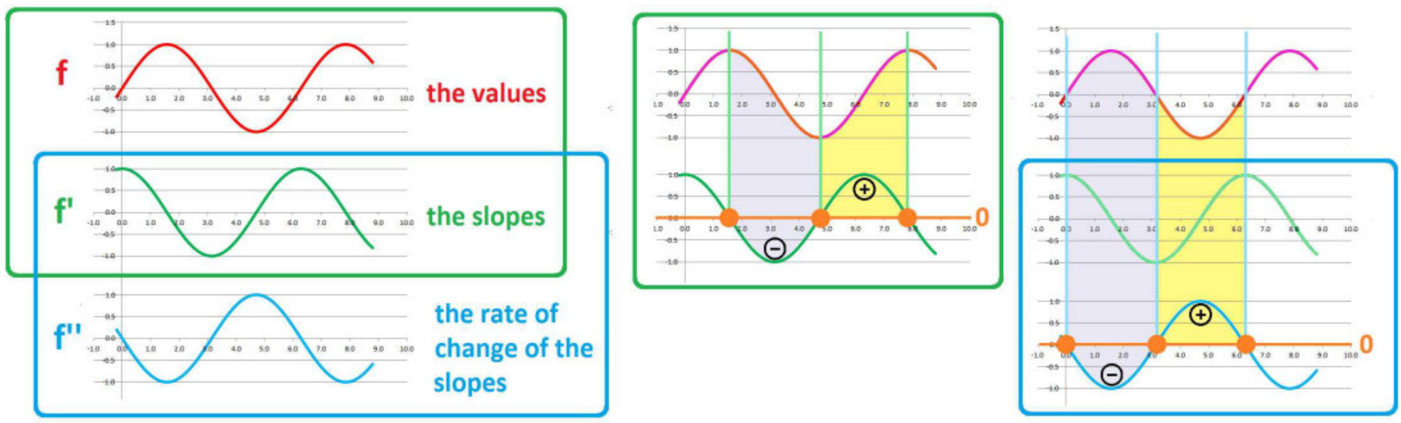


Specifically, we see:

1. Decreasing slopes $\iff$ tangents rotate clockwise.
2. Increasing slopes $\iff$ tangents rotate counter-clockwise.

This matches our convention from trigonometry that counter-clockwise is the positive direction for rotations.

Even though we typically have functions with the nth derivative for each positive integer n, only the first two reveal something *visible* about the graph of the original function. Below, we "derive" the graph of the derivative f'' for that of the function by looking at the slopes of f; then we "derive" the graph of the derivative f'' of f' for that of f' by looking at its slopes of f':



What is new? The relation between f and f''. Above we compare the following pairs:

- Old: We compare the *shapes* of the patches of the graph of the function f to the sign of the *values* of the first derivative f'.
- New: We compare the *shapes* of the patches of the graph of the function f to the sign of the *values* of the second derivative f''.

There are three main levels of analysis of a function:

- *Analysis at level* 0: the values of f. We ask, how large? The findings are about the values, x- and y-intercepts, asymptotes and other large-scale behavior, periodicity, etc.
- *Analysis at level* 1: the slopes of f. We ask, up or down? The findings are about the angles, increasing/decreasing behavior, critical points, etc.
- *Analysis at level* 2: the rate of change of the slopes of f. We ask, concave up or down? The findings are about the change of steepness, concavity, telling a maximum from a minimum, etc.

We can go on and continue to discover more and more subtle but less and less significant properties of the function.

Example 4.6.11: motion

This three-level analysis also applies to our study of motion, as follows.

The derivative of the velocity and, therefore, the second derivative of the position, is called the *acceleration.* The concept allows one to add another level of analysis of motion:

- Analysis at level 0: the location, where?
- Analysis at level 1: the velocity, how fast? forward or back?
- Analysis at level 2: the acceleration, how large is the force?

Suppose t is time and y is the vertical dimension, the height. Now we turn to the specific case of *free fall.* These are the initial conditions:

- y_0 is the initial height, $y_0 = y\Big|_{t=0}$.
- v_y is the initial vertical component of velocity, $\frac{dy}{dt}\Big|_{t=0}$.

Then, we have:

$$y \;= y_0 + v_y t - \frac{1}{2}gt^2 \quad\Longrightarrow\quad \frac{dy}{dt} \;= v_y \quad -gt \quad\Longrightarrow\quad \frac{d^2y}{dt^2} \;= -g\,.$$

Now, from the point of the physics of the situation, the derivation should go in the opposite direction:

- When there is no force, the velocity is constant.
- When the force is constant, the velocity is linear on time, etc.

However, at this point we are still unable to answer these questions:

- How do we know that only the derivatives of constant functions and none others are zero?
- How do we know that only the derivatives of linear functions and none others are constant?
- How do we know that only the derivatives of quadratic functions and none others are linear?

This reversed process is called *antidifferentiation.*

So far, we cannot justify even this simplest conclusion:

$$f' = 0 \Longrightarrow\; f = c, \quad \text{for some real number } c\,.$$

We will study these and related questions in Chapter 5.

4.7. How to differentiate relations: implicitly

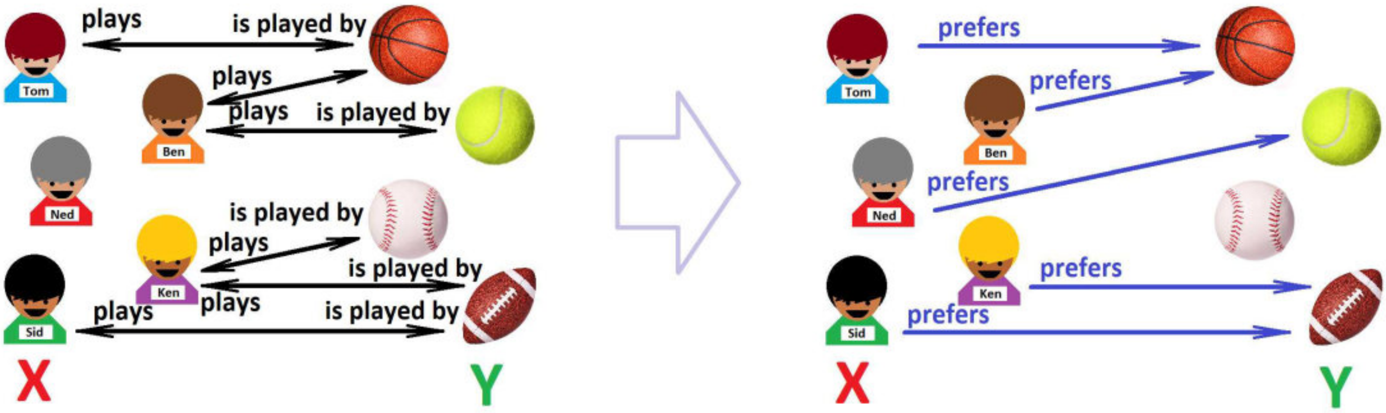
It seems that we can find the tangent line to any curve! Yes, with a caveat: This curve is the *graph of a function.* How do we find the tangent to a circle? It is given by a relation:

$$x^2 + y^2 = 1\,.$$

Recall that a *function* is a special kind of relation. These two are the two main constructs calculus relies on:

- A relation: "What sports have been played by what boys today?"
- A function: "Which sport does the boy *prefer* to play?"

The difference is that in the latter case everyone has a preference and exactly one:



We can differentiate functions; can we differentiate *relations*?

Recall (from Volume 1, Chapter 1PC-2) that relations are represented by equations.

Usually, equations for x are to be solved:

$$\underbrace{x^2}_{x \text{ is a number}} - \underbrace{1}_{\text{a number}} = 0\,. \quad \text{Now find a particular number } x \text{ that satisfies the equation.}$$

The equations we are interested in are equations of x's and y's, such as:

$$\underbrace{x^2}_{x \text{ is a number}} + \underbrace{y^2}_{y \text{ is a function}} = 0\,. \quad \text{Now find a particular function } y = y(x) \text{ that satisfies the equation.}$$

In other words, after the substitution, the equation should be true for *all* x, i.e., the functions on the left and on the right are identical ($x + x = 2x$, etc.).

The equation *implicitly* defines this function. As we have done in the past, we could make the function $y = y(x)$ *explicit* by solving the equation for y:

$$y = \sqrt{1 - x^2} \quad \text{or} \quad y = -\sqrt{1 - x^2}\,.$$

But we only want the tangent, i.e., the rate of change of this function! We will leave y *unspecified*.

We will rely on the following fact:

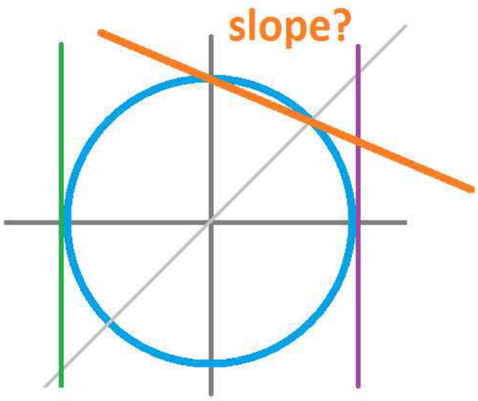
▶ If two functions are identical, for all nodes x of a partition, then so are their difference quotients for all secondary nodes c:

$$\boxed{f(x) = g(x) \ \text{ for all } x \implies \frac{\Delta f}{\Delta x}(c) = \frac{\Delta g}{\Delta x}(c) \ \text{ for all } c}$$

Example 4.7.1: circle

We start with finding the *secant line* through the two points on the circle of radius 1 centered at 0:

$$\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right) \ \text{and any other point } (x, y)\,.$$



Typically, a curve has been the graph of a function $y = x^2$, $y = \sin x$, etc., given explicitly. This time the equation is:

$$x^2 + y^2 = 1\,.$$

To find the slope of the secant line, we need the difference quotient of the function. But there is no function! Not explicit anyway.

The idea is to consider the above equation as a relation between the two variables. In fact, we think of $y = y(x)$ as a function of x, i.e.:

$$x^2 + y(x)^2 = 1\,.$$

We will also assume:

- The two x-values x_0 and $x_1 = \dfrac{\sqrt{2}}{2}$ are nodes of a partition of the x-axis.
- The two y-values y_0 and $y_1 = \dfrac{\sqrt{2}}{2}$ are nodes of a partition of the y-axis.

We apply the *Chain Rule* to both sides of the equation:

$$\begin{aligned}
\frac{\Delta}{\Delta x}\left(x^2 + y^2\right) &= \frac{\Delta}{\Delta x}(1) && \Longrightarrow\\
\frac{\Delta}{\Delta x}x^2 + \frac{\Delta}{\Delta x}y^2 &= 0 && \Longrightarrow\\
(x_0 + x_1) + (y_0 + y_1)\frac{\Delta y}{\Delta x} &= 0 && \Longrightarrow\\
\frac{\Delta y}{\Delta x} &= -\frac{x_0 + x_1}{y_0 + y_1} && \text{for } y_0 + y_1 \neq 0\,.
\end{aligned}$$

We have found a formula for the difference quotient but it is still implicit – because we don't have a formula for $y = y(x)$. Fortunately, we don't need the whole function, just those two points on its graph. We substitute these into the formula above to find:

$$\frac{\Delta y}{\Delta x} = -\frac{y_0 + \frac{\sqrt{2}}{2}}{x_0 + \frac{\sqrt{2}}{2}}\,.$$

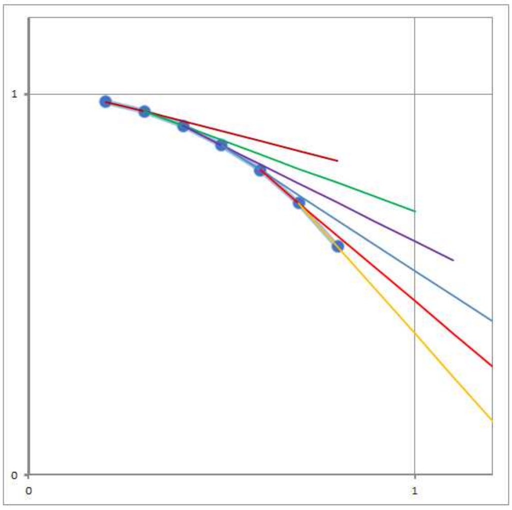
In particular, for the point $(x_0, y_0) = (0, 1)$, the slope is

$$m = -\frac{\sqrt{2}}{1 + \sqrt{2}}\,.$$

Then, from the *point-slope formula* we obtain the answer:

$$y - \frac{\sqrt{2}}{2} = -\frac{\sqrt{2}}{1 + \sqrt{2}}\left(x - \frac{\sqrt{2}}{2}\right)\,.$$

We can automate this formula and find more secant lines with a spreadsheet:



What about the *derivative*? We will rely on the following fact:

▶ If the values of two functions are equal for all x, then so are the values of their derivatives:

$$\boxed{f(x) = g(x) \text{ for all } x \implies f'(x) = g'(x) \text{ for all } x}$$

We can put it simply as this:

▶ If two functions are identical, then so are their derivatives; i.e.,

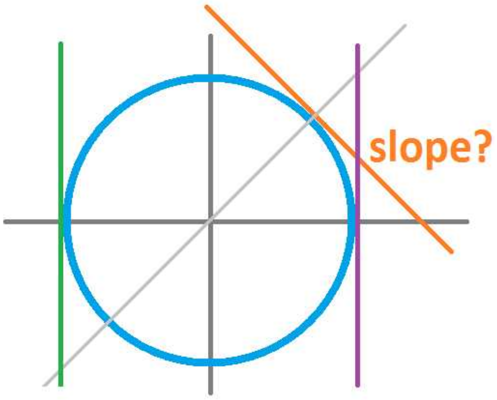
$$\boxed{f = g \implies f' = g'}$$

Differentiating an equation of functions and finding the derivative of a function defined by this equation is called *implicit differentiation.*

Let's consider two examples of how this idea may help us with finding tangents to implicit curves.

Example 4.7.2: circle

Find the *tangent line* for the circle of radius 1 centered at 0 at the point $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$.



Typically, a curve has been the graph of a function $y = x^2$, $y = \sin x$, etc., given explicitly. This time the equation is:

$$x^2 + y^2 = 1\,.$$

To find the slope of the tangent line, we need the derivative, but there is no function to differentiate!

Our approach is to *differentiate the equation* above as a relation between the two variables. As we

differentiate, we think of $y = y(x)$ as a function of x. Therefore, $y^2 = y(x)^2$ is the composition:

$$\begin{array}{lccccc} \text{variables:} & x & \xrightarrow{y} & y & \xrightarrow{y^2} & z \\ \text{derivatives:} & & \dfrac{dy}{dx} & & 2y & \end{array}$$

We differentiate the equation (the identity of functions):

$$x^2 + y^2 = 1\,.$$

This is the result, via the Chain Rule:

$$\begin{array}{rll} \dfrac{d}{dx}\left(x^2+y^2\right) & = \dfrac{d}{dx}(1) & \Longrightarrow \\ \dfrac{d}{dx}x^2 + \dfrac{d}{dx}y^2 & = 0 & \Longrightarrow \\ 2x + 2y\dfrac{dy}{dx} & = 0 & \Longrightarrow \\ \dfrac{dy}{dx} & = -\dfrac{x}{y} & \text{for } y \neq 0\,. \end{array}$$

We have found a formula for the derivative, but it is still implicit – because we don't have a formula for $y = y(x)$. Fortunately, we don't need the whole function, just a single point on its graph:

$$x = \frac{\sqrt{2}}{2},\ y = \frac{\sqrt{2}}{2}\,.$$

We substitute these into the formula above to find:

$$\left.\frac{dy}{dx}\right|_{x=\frac{\sqrt{2}}{2},\ y=\frac{\sqrt{2}}{2}} = -\left.\frac{x}{y}\right|_{x=\frac{\sqrt{2}}{2},\ y=\frac{\sqrt{2}}{2}} = -1\,.$$

Finally, from the point-slope formula we obtain the answer:

$$y - \frac{\sqrt{2}}{2} = -1\left(x - \frac{\sqrt{2}}{2}\right)\,.$$

We could use the explicit formula $y = \sqrt{1-x^2}$ with the same result:

$$\frac{dy}{dx} \overset{\text{CR}}{=\!=} \frac{-2x}{2\sqrt{1-x^2}} = -\frac{x}{1-x^2}\,,$$

after we substitute $x = \dfrac{\sqrt{2}}{2}$. However, the implicit approach required a single differentiation ($(y^2)' = 2y$) in comparison. Also, the formula is only explicit for the upper half of the circle. For a point below the x-axis, we'd need to start over and use the other formula, $y = -\sqrt{1-x^2}$.

No matter what the approach is, the derivative $\dfrac{dy}{dx}$ is undefined at $x = \pm 1$ because the denominator is 0. How do we find the tangent? From the implicit formula, we can proceed in two directions:

$$x^2 + y^2 = 1 \quad \begin{array}{l} \nearrow\ y \text{ depends on } x \\ \searrow\ x \text{ depends on } y \end{array}$$

Then, we can try implicit differentiation of the same equation – but with respect to y this time. The computation is very similar, and the result is:

$$\frac{dx}{dy} = -\frac{y}{x}\,.$$

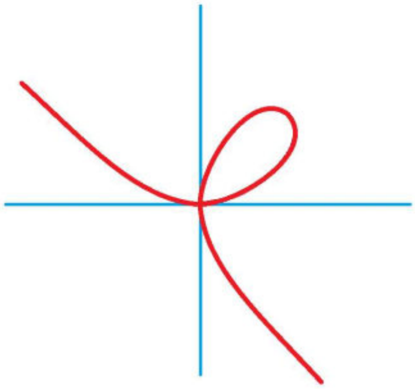
The formula *is* defined for $y = 0$, at the points $(-1, 0),\ (1, 0)$. Then, $\dfrac{dx}{dy} = 0$ at these points. Therefore, the tangent line is $x - 1 = 0(y - 0)$, or $x = 1$.

Example 4.7.3: Folium of Descartes

This curve is given by the equation

$$x^3 + y^3 = 6xy$$

plotted below:



Let's find some tangents. We differentiate the equation as before:

$$\frac{d}{dx}\left(x^3 + y^3\right) = \frac{d}{dx}(6xy) .$$

While using the *Chain Rule*, we notice that every time we see y, the factor $\frac{dy}{dx}$ also appears:

$$\begin{aligned} \frac{d}{dx}(x^3) + \frac{d}{dx}(y^3) &= 6\frac{d}{dx}(xy) && \Longrightarrow \\ 3x^2 + 3y^2 \cdot \frac{dy}{dx} &= 6\left(y + x\frac{dy}{dx}\right) . \end{aligned}$$

Solve for $\frac{dy}{dx}$:

$$\begin{aligned} 3x^2 + 3y^2\frac{dy}{dx} &= 6y + 6x\frac{dy}{dx} && \Longrightarrow \\ (3y^2 - 6x)\frac{dy}{dx} &= 6y - 3x^2 && \Longrightarrow \\ \frac{dy}{dx} &= \frac{6y - 3x^2}{3y^2 - 6x} . && \text{As long as this isn't } (0,0). \end{aligned}$$

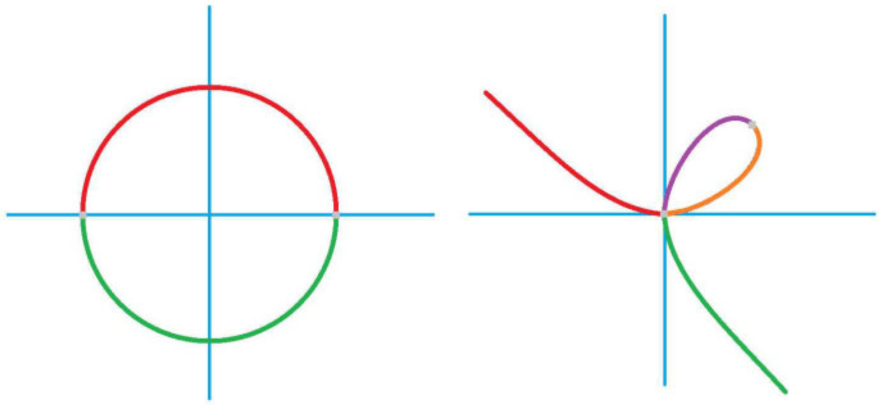
The end result is: If we know the location (x, y), we know the slope of the tangent at that point. For example, at the tip of the curve we have $x = y$. Therefore, the slope is $\frac{dy}{dx} = -1$.

Exercise 4.7.4

Find the tangents at the origin.

In both examples, we chose calculus over algebra, and the choice led – via the Chain Rule – to simpler computations.

Note that in either example, we can cut the curve into pieces each of which is the graph of a function:



Exercise 4.7.5

Find other ways to do it.

In summary, if there is a curve given by a relation, we can declare one of the variable independent and the other dependent and find an implicit representation of the derivative of the latter over the former, which is the corresponding slope.

Now, implicit differentiation also helps with situations when two or more quantities depend on each other via a relation as well on a common independent variable, usually the time. If we differentiate this relation with respect to time, we get a relation among these quantities and their derivatives. The result is called the *related rates*.

Example 4.7.6: air balloon

Suppose we have an air balloon, spherical in shape. Air is pumped in it and at a certain moment of time it was recorded to be at the rate of $5^{\text{in}^3}/_{\text{sec}}$. How fast does the radius grow?

Step 1 in word problems is to introduce variables. Let

- t be time,
- V be the volume, and
- r be the radius.

We can also name that specific moment of time, say, t_0.

Next, V depends on t. Even though this dependence will remain unknown, we do know its derivative at that moment t_0. Then,

$$\left.\frac{dV}{dt}\right|_{t=t_0} = 5\,,$$

according to the condition. Furthermore, this is a sphere, so the known formula for its volume is the following:

$$V = \frac{4}{3}\pi r^3\,.$$

Here we see that V also depends on r; altogether, this is the dependencies we face:

$$\begin{array}{ccc} t & \to & r \\ & \searrow & \downarrow \\ & & V \end{array}$$

We *could* reverse the last arrow by finding the inverse:

$$r = \sqrt[3]{\frac{3}{4\pi}V}\,,$$

and deal with this complex algebra. Instead, we *differentiate the equation itself*. With respect to t, of course! Thus, if two variables are related (via an equation), then so are their derivatives, i.e., the

rates of change. Hence, "related rates".

Keeping in mind that both V and r are functions of time, we differentiate the relation with respect to t:

$$V = \frac{4}{3}\pi r^3 \,.$$

The left-hand side is very simple:

$$\frac{d}{dt}V = \frac{dV}{dt}\,,$$

but in the right-hand side, $r(t)^3$ is a composition! We differentiate by the Chain Rule:

$$\frac{d}{dt}\left(\frac{4}{3}\pi r^3\right) = \frac{4}{3}\pi \cdot 3r^2\frac{dr}{dt}\,.$$

Thus, we have:

$$\frac{dV}{dt} = 4\pi \cdot 3r^2\frac{dr}{dt}\,.$$

Recall that the rate of change of V is 5 (at $t = t_0$), so we have:

$$5 = 4\pi r^2\frac{dr}{dt} \implies \frac{dr}{dt} = \frac{5}{4\pi r^2}\,.$$

Therefore, the rates of growth of r for the cases $r = 1,\ r = 2,\ r = 3$ are:

$$\begin{aligned} r = 1: &\quad \frac{dr}{dt} = \frac{5}{4\pi} \\ r = 2: &\quad \frac{dr}{dt} = \frac{5}{16\pi} \\ r = 3: &\quad \frac{dr}{dt} = \frac{5}{36\pi} \end{aligned}$$

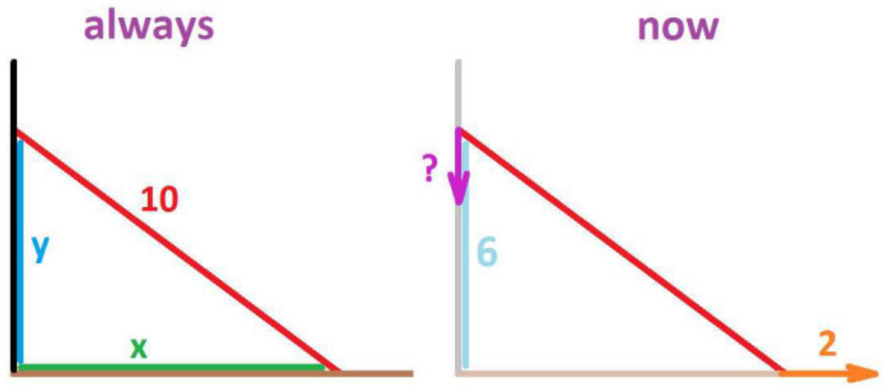
So, the effect on the radius of pumping in the air is smaller for larger radii.

Exercise 4.7.7

How fast is the surface area growing?

Example 4.7.8: sliding ladder

Suppose a 10-foot ladder stands against the wall and its bottom is sliding away from the wall. How fast is the top moving when it is 6 ft from the floor and the bottom is sliding at 2 ft/sec?



Introduce variables:

- x the distance of the bottom from the wall,
- y the distance of the top from the floor, both functions of
- t the time.

Let's denote the moment of time t_0.

We translate the information, as well as the question, about the variables into equations:

	quantities:	functions:
always	$x^2+y^2=10^2$	$(x(t))^2+(y(t))^2=10^2$ for all t
now	$x=?$	$x(t_0)=?$
now	$\dfrac{dx}{dt}=2$	$x'(t_0)=2$
now	$y=6$	$y(t_0)=6$
now	$\dfrac{dy}{dt}=?$	$y'(t_0)=?$

That's a purely mathematical problem to be solved.

We differentiate the equation with respect to the independent variable, t:

$$\begin{array}{rll}\dfrac{d}{dt}\left(x^2+y^2\right) & =\dfrac{d}{dt}(100) & \text{We apply CR twice.}\\ 2x\dfrac{dx}{dt}+2y\dfrac{dy}{dt} & =0 & \text{Solve for } \dfrac{dy}{dt}.\\ \dfrac{dy}{dt} & =-\dfrac{x}{y}\dfrac{dx}{dt} & \end{array}$$

This is a relation of *four* functions!

At the specific moment $t=t_0$, we find $x=8$ from $x^2+y^2=100$. From this and the fact that $y=6$, we conclude:

$$\left.\frac{dy}{dt}\right|_{t=t_0}=-\frac{8}{6}2=-\frac{8}{3}\,.$$

Exercise 4.7.9

Solve the problem for the moment when the ladder is upright.

Exercise 4.7.10

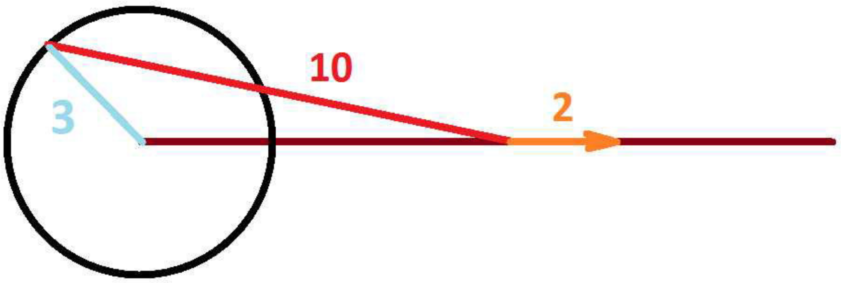
Solve the problem for the moment when the ladder hits the floor.

Exercise 4.7.11

Plot $y=y(t)$ based on this information.

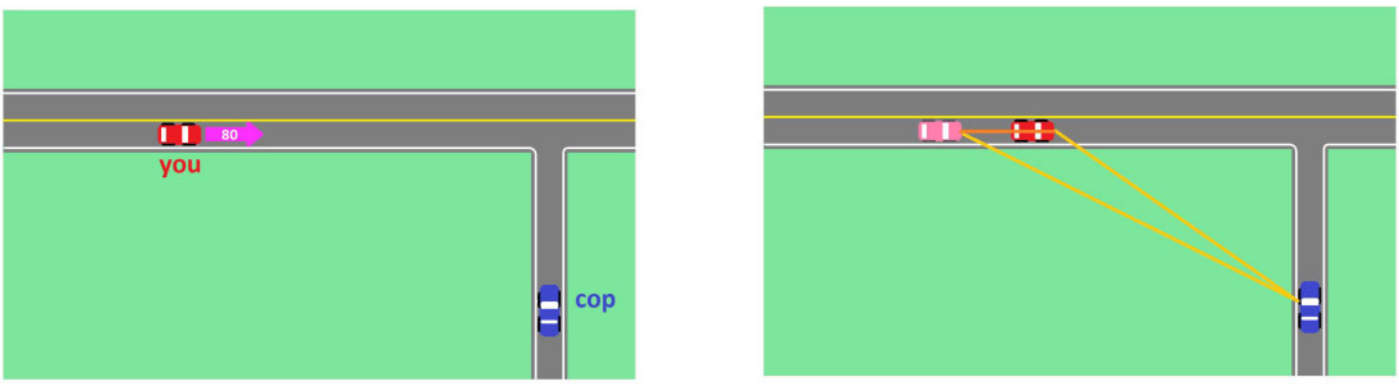
Exercise 4.7.12

Set up and solve a similar problem for the crankshaft:



4.8. Related rates: radar gun

Problem: Suppose you are driving at a speed of 80 mph when you see a police car positioned 40 feet off the road. What is the radar gun's reading?



First, how does the radar gun work? In fact, how does a *radar* work? A signal is sent, it bounces off an object, and, when it comes back, the time lapse is recorded. Then, the distance to the object is computed as follows:

$$S = \underbrace{\text{signal's speed}}_{\text{known}} \cdot \underbrace{\text{time passed}}_{\text{measured}}$$

A radar gun does this *twice*.

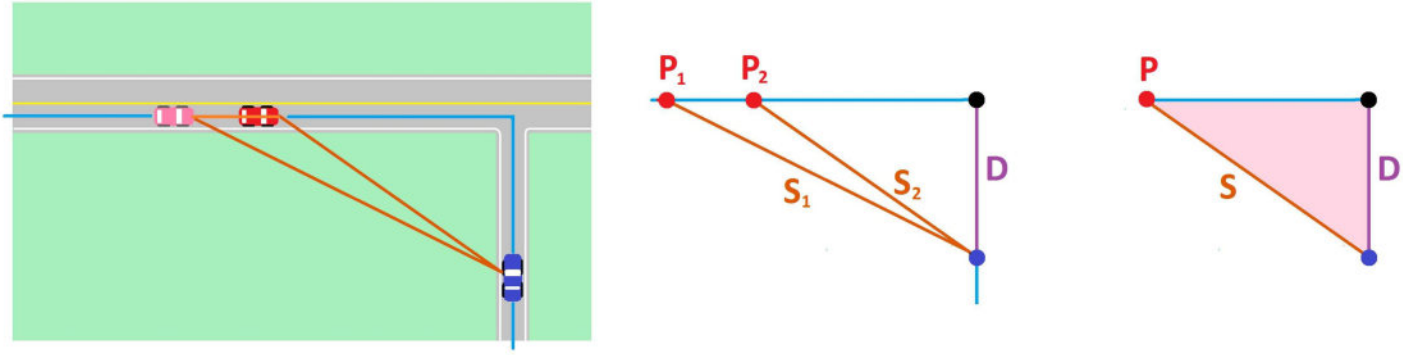
A signal is sent, it comes back, the time is measured and the distance computed. Then the second time:

- $S_1 = \text{speed} \cdot \text{time}$, at time $t = t_1$,
- $S_2 = \text{speed} \cdot \text{time}$, at time $t = t_2$.

Then, the reading is computed as the difference quotient of the distance between the signal source (the police car) and the target (your car):

$$\text{estimated speed} \;=\; \frac{\text{change of distance}}{\text{time between signals}}\,.$$

No radar gun can do better than that!



If the signal is continued to be emitted, the distance is a function $S = S(t)$ known at a sequence of moments of time $t_0, t_1, \ldots$. Then the radar computes the difference quotient of S:

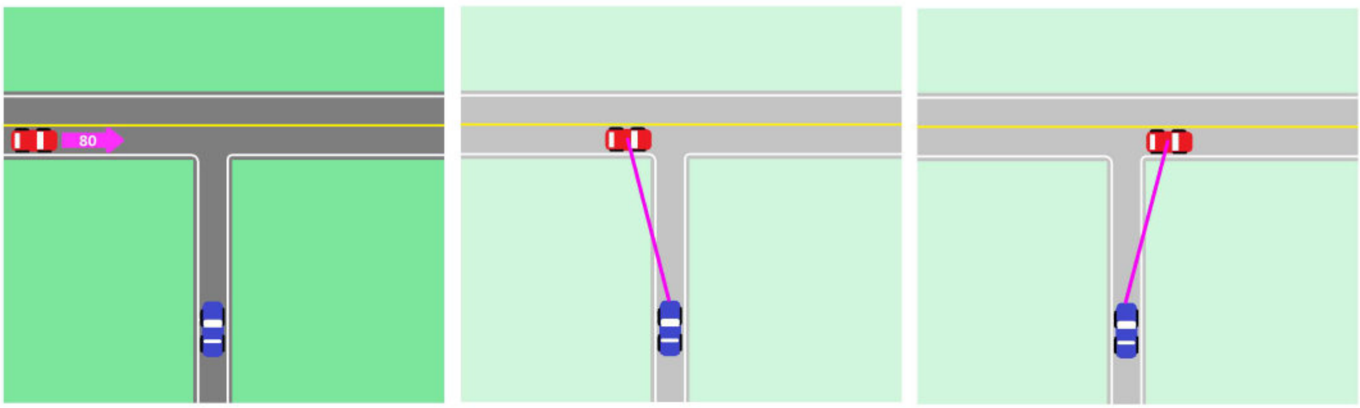
$$\text{Radar reading} \;=\frac{\Delta S}{\Delta t},$$

where

- $\Delta S = S_{n+1} - S_n$ is the change of the distance between the two cars,
- $\Delta t = h = t_{n+1} - t_n$ is the time between signals.

Now the question, is the reading of the radar gun 80 m/h?

To get an idea of what can happen, consider this extreme example: What if you are just passing in front of the police car, like this?

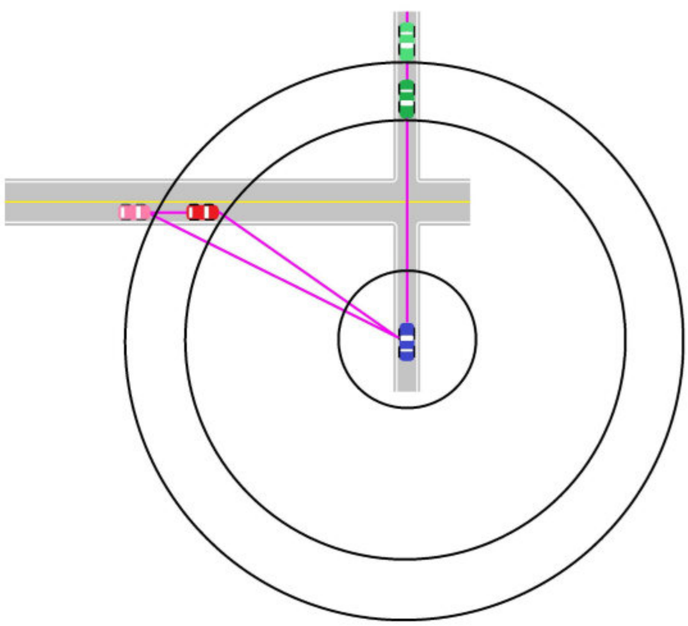


It is conceivable that at time t_1 your car is the same distance from the intersection as it is past the intersection at time t_2. Then

$$\Delta S = 0 \implies \frac{\Delta S}{\Delta t} = 0\,.$$

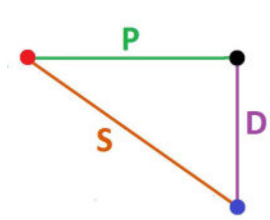
The reading is off by a lot!

We can see the source of the error below:



The radar gun tells only how fast the other car is going through these concentric circles, not its actual speed. The reason for the error is that the direction of the red car is askew. Meanwhile the reading for the green car would be accurate because its direction is straight at the radar.

Let's start over. We have a right triangle that changes its shape with time:



These are the quantities:

- S, the distance between the police car to yours
- P, the distance between your car to the intersection
- t, the time, the independent variable
- $D = 40$, distance from the police car to the road

Since 80 ml/h is your speed, we have $\frac{dP}{dt} = 80$. That's what the radar gun is meant to detect. But what it does measure in reality is $\frac{dS}{dt}$!

This is what we need to find out:

▶ How good an approximation of the real velocity $\frac{dP}{dt}$ is the perceived velocity $\frac{dS}{dt}$?

Our spreadsheet contains a column of locations P of your car (distances to the intersection) found by means of the familiar formula

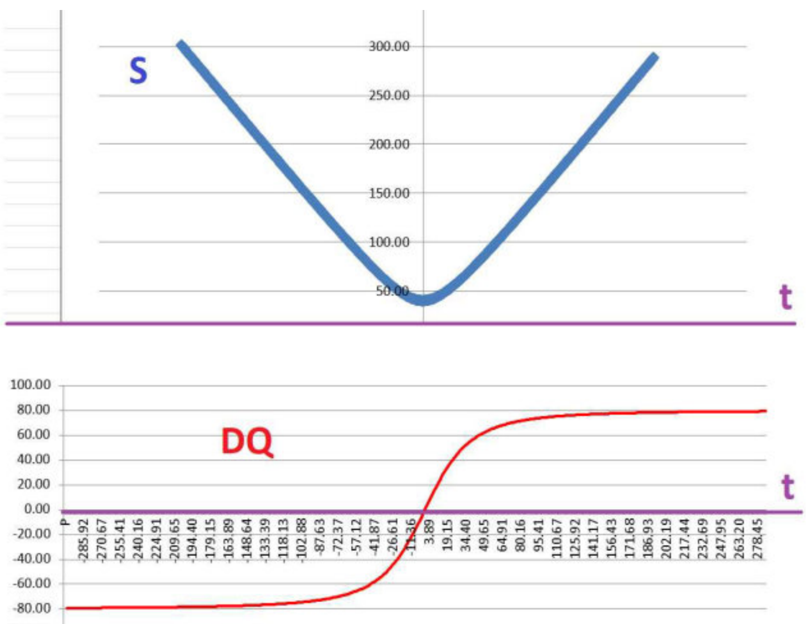
$$P_{n+1} = P_n + 80h\,,$$

where h is the increment of time. The next column is for the distance S to the police car (plotted first), found via the *Pythagorean Theorem*:

$$P^2 + D^2 = S^2\,.$$

The third column is the difference quotient of S (plotted second):

P	S	DQ
-300.00	302.65	
-298.83	301.49	-79.30
-297.65	300.33	-79.29
-296.48	299.17	-79.28
-295.31	298.00	-79.28
-294.13	296.84	-79.27
-292.96	295.68	-79.27
-291.79	294.52	-79.26
-290.61	293.35	-79.26
-289.44	292.19	-79.25
-288.27	291.03	-79.24
-287.09	289.87	-79.24
-285.92	288.70	-79.23
-284.75	287.54	-79.23
-283.57	286.38	-79.22
-282.40	285.22	-79.21
-281.23	284.06	-79.21
-280.05	282.90	-79.20
-278.88	281.73	-79.19
-277.71	280.57	-79.19
-276.53	279.41	-79.18
-275.36	278.25	-79.17
-274.19	277.09	-79.17
-273.01	275.93	-79.16
-271.84	274.77	-79.15
-270.67	273.61	-79.14
-269.49	272.45	-79.14
-268.32	271.29	-79.13



As we can see, the approximation is the best when the red car is away from the intersection. But, within 75 feet from the intersection, the reading will be less than 70 mph!

From this approximation, we move to a more precise analysis.

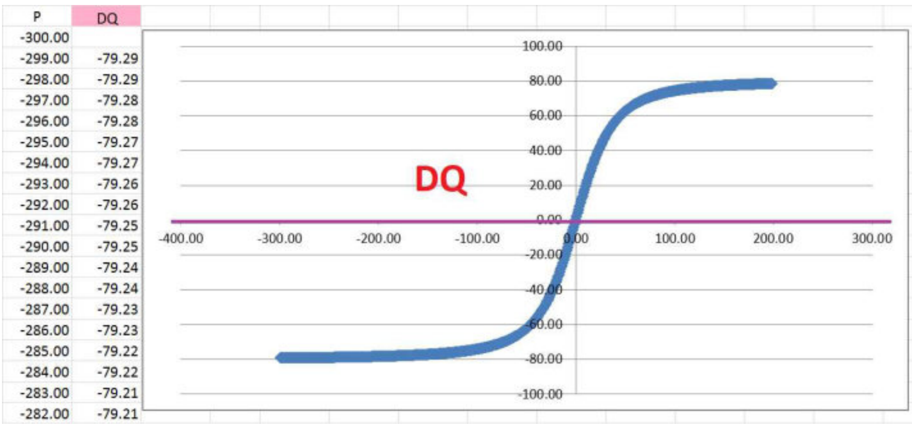
We have a relation between the two variables via the Pythagorean Theorem:

$$P^2 + D^2 = S^2 \ \leftarrow \ \text{These aren't numbers, but variables, i.e., functions.}$$

This equation connects P and S, but not $\frac{dP}{dt}$ and $\frac{dS}{dt}$ yet. We differentiate the equation with respect to t:

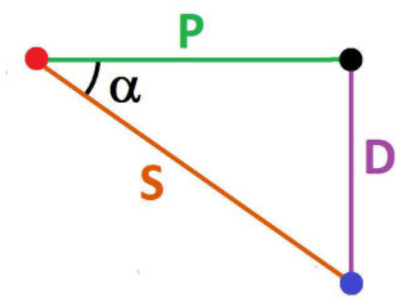
$$\begin{aligned}
\frac{d}{dt}\left(P^2+D^2\right) &= \frac{d}{dt}\left(S^2\right) \implies \\
2P\cdot\frac{dP}{dt}+2D\underbrace{\frac{dD}{dt}}_{=0} &= 2S\cdot\frac{dS}{dt} \implies \\
P\cdot\frac{dP}{dt} &= S\cdot\frac{dS}{dt} \implies \\
\frac{dS}{dt} &= \frac{P}{S}\frac{dP}{dt}\,.
\end{aligned}$$

Thus, we finally have a relation between these functions. We plot this function below, to confirm our earlier conclusions:



P	DQ
-300.00	
-299.00	-79.29
-298.00	-79.29
-297.00	-79.28
-296.00	-79.28
-295.00	-79.27
-294.00	-79.27
-293.00	-79.26
-292.00	-79.26
-291.00	-79.25
-290.00	-79.25
-289.00	-79.24
-288.00	-79.24
-287.00	-79.23
-286.00	-79.23
-285.00	-79.22
-284.00	-79.22
-283.00	-79.21
-282.00	-79.21

Another way to approach the problem is the POV of the driver. Let α be the angle between the road ahead of you and the direction to the police car:



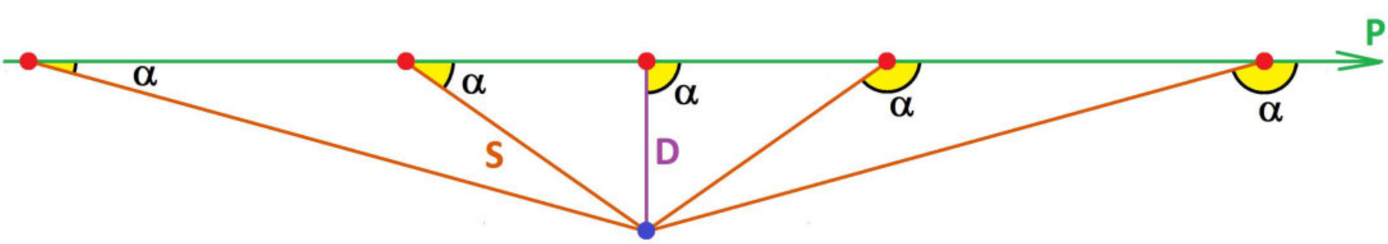
Then:

$$\cos\alpha = \frac{P}{S}\,.$$

In other words, we have:

$$\text{Radar reading} \;=\; \text{actual speed} \;\cdot\; \cos\alpha\,.$$

This is how does α change as you drive:



These are the conclusions:

- Early, α is close to 0, so $\cos\alpha$ close to 1, and, therefore, $\frac{dS}{dt}$ is close to 80.

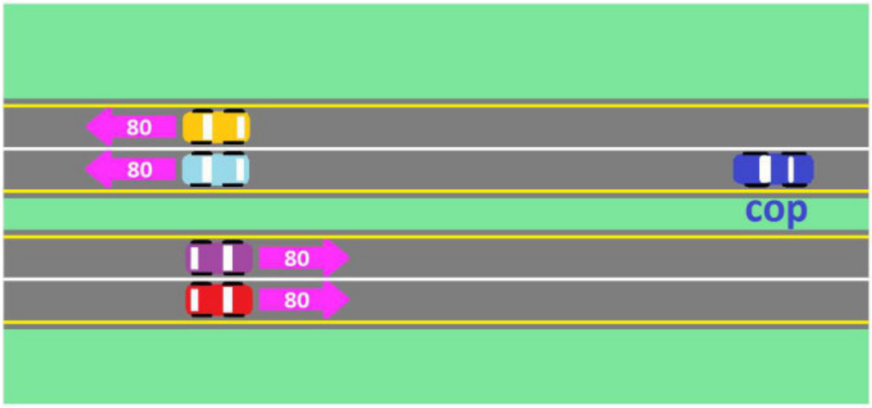
- Then, as α increases, $\cos\alpha$ decreases toward 0, and so does $\frac{dS}{dt}$.
- In the middle, we have $\alpha = \frac{\pi}{2}$, $\cos\alpha = 0$, $\frac{dS}{dt} = 0$.
- As α passes $\frac{\pi}{2}$, $\cos\alpha$ decreases to negative values, and so does $\frac{dS}{dt}$.
- Later, α approaches π, and $\cos\alpha$ approaches 1, and, therefore, $\frac{dS}{dt}$ approaches 80.

Conclusion: The radar gun always *under*estimates your speed:

$$\left|\frac{dS}{dt}\right| < 80\,.$$

Unless, the police car is *on* the road!

In that case, what can you do to “improve” the reading? What do you want α to be – as large as possible!



Exercise 4.8.1

Consider the POV of the policeman and the other angle β of the triangle.

Exercise 4.8.2

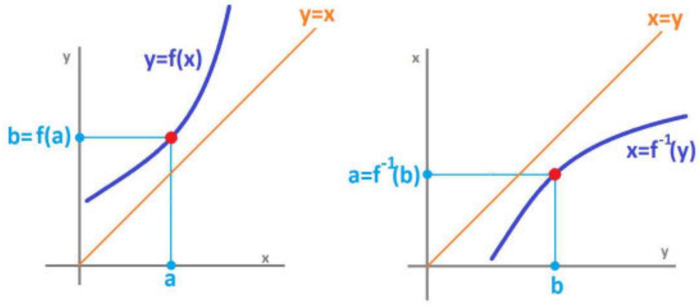
Show that, with preparation, it is possible for the policeman to find the exact value of the other car's speed.

4.9. The derivative of the inverse function

The list of functions we cannot differentiate yet remains long:

$$\sqrt{x},\ \ln x,\ \arcsin x, \ldots$$

What do these have in common? They are the inverses of functions with known derivatives!



Let's recall (from Volume 1, Chapter 1PC-3) that for a given one-to-one and onto function $y = f(x)$, its *inverse* is the function, $x = f^{-1}(y)$, that satisfies

$$f^{-1}(y) = x \iff f(x) = y\,.$$

The idea is that a function and its inverse represent the *same relation*:

- x and y are related when $y = F(x)$, or
- x and y are related when $x = F^{-1}(y)$.

For example, these are pairs of functions inverse to each other:

$$\begin{array}{rcll} y = x + 2 & \text{vs.} & x = y - 2 & \\ y = 3x & \text{vs.} & x = \dfrac{1}{3}y & \\ y = x^2 & \text{vs.} & x = \sqrt{y} & \text{for } x, y \geq 0 \\ y = e^x & \text{vs.} & x = \ln y & \text{for } y > 0 \end{array}$$

Can we express *the derivative of the inverse of a function in terms of the derivative of the function*?

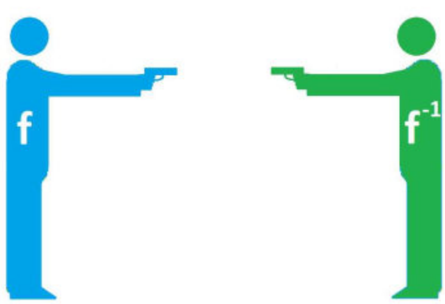
We will utilize the following, algebraic, definition of the inverse (seen in Volume 1, Chapter 1PC-3):

$$\boxed{\begin{array}{lll} f^{-1}(f(x)) & = x & \text{for all } x \\ f\left(f^{-1}(y)\right) & = y & \text{for all } y \end{array}}$$

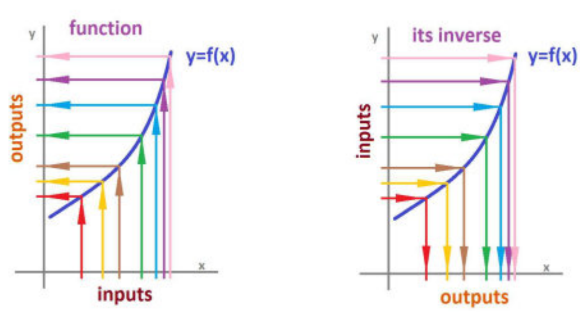
Here is a flowchart representation of this idea:

$$\begin{array}{llllllll} x & \to & \boxed{\quad f \quad} & \to\ y\ \to & \boxed{\quad f^{-1} \quad} & \to & x & \text{(same } x\text{)} \\ y & \to & \boxed{\quad f^{-1} \quad} & \to\ x\ \to & \boxed{\quad f \quad} & \to & y & \text{(same } y\text{)} \end{array}$$

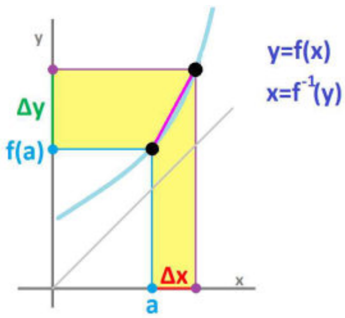
In other words, when combined they kill each other:



Let's recall that the inverse "undoes" the effect of the function:



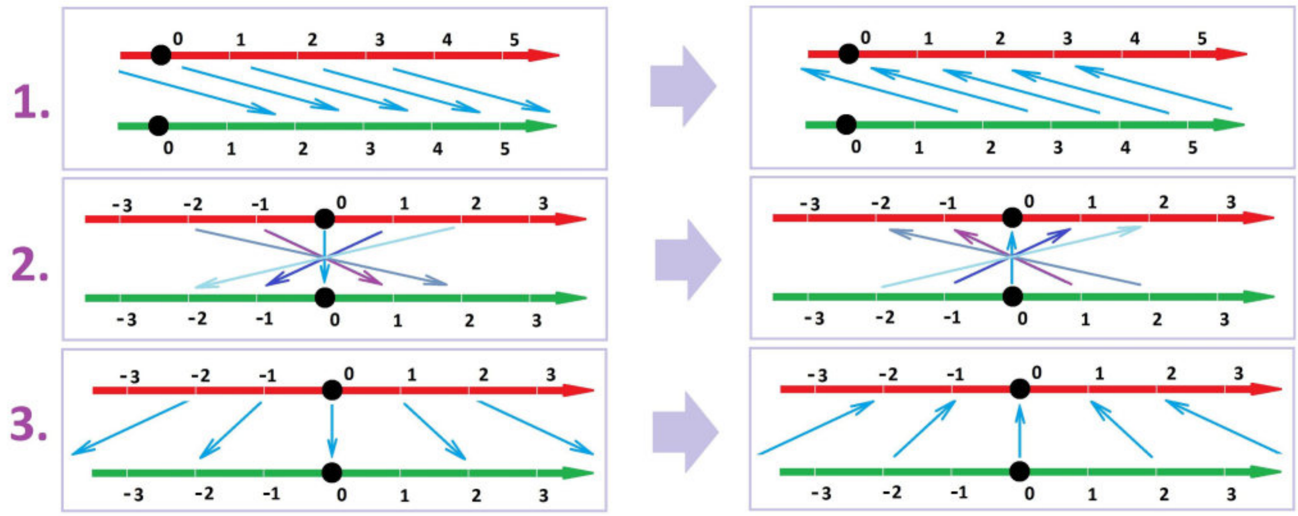
We can see how the difference of x becomes the difference of y, under f and f^{-1}:



There is no other relation!

Example 4.9.1: transformations

The idea that the inverse reverses the effect of the function applies especially well to *transformations*:



Now, what is the meaning of the difference quotient (and the derivative) of a transformation? It is its stretch ratio. Now, this is common sense:

► If f *stretches* the x-axis at the rate of k (at $x = a$), then f^{-1} *shrinks* the y-axis at the rate of k (at $b = f(a)$).

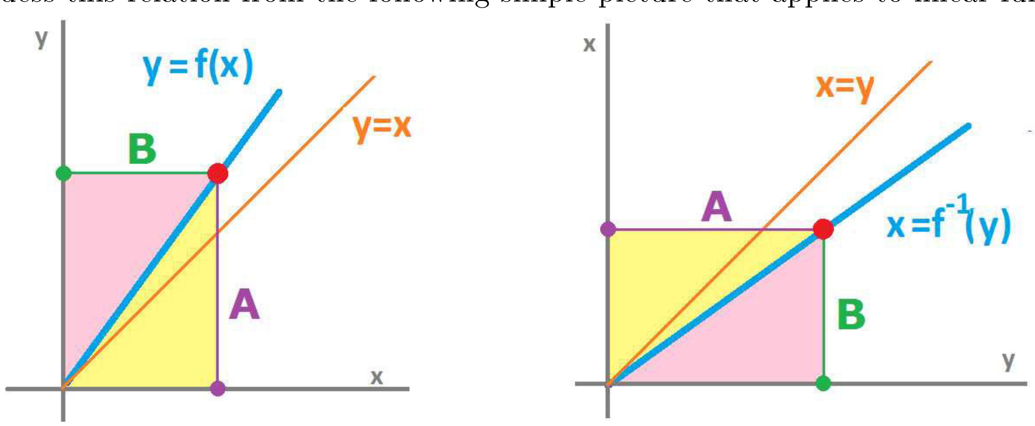
So, we have a match:

$$k \ \text{ for } f \ \text{ and } \frac{1}{k} \ \text{ for } \ f^{-1}\,.$$

It's the *reciprocal*!

Example 4.9.2: linear functions

We can also guess this relation from the following simple picture that applies to linear functions:



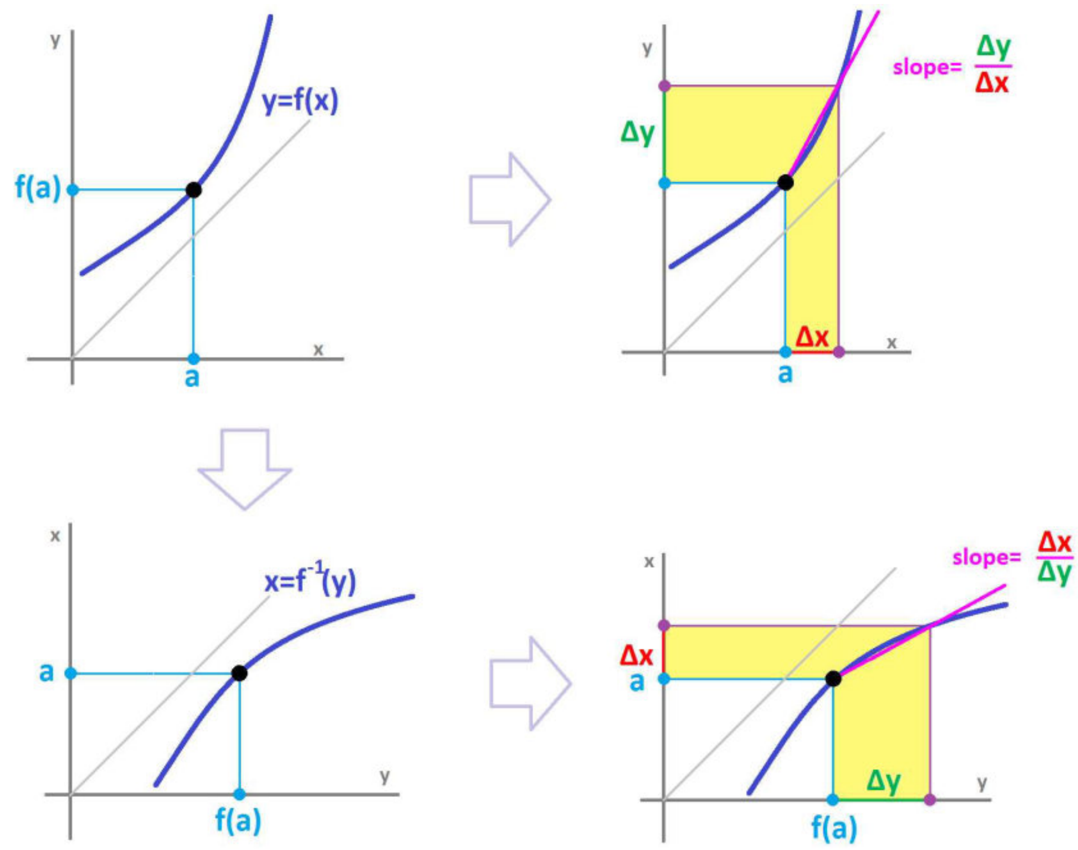
As the xy-plane is flipped about the diagonal, this is what happens:

$$\text{slope of } f = \frac{\text{change of } y}{\text{change of } x} = \frac{A}{B} = \frac{1}{B/A} = \frac{1}{\text{slope of } f^{-1}}\,.$$

We have conjectured a formula:

$$\frac{\Delta f^{-1}}{\Delta y} = \frac{1}{\frac{\Delta f}{\Delta x}}\,.$$

Even though the derivatives aren't fractions, the difference quotients, i.e., the slopes of the secant lines, are:



This analysis proves the following:

Theorem 4.9.3: Difference Quotient of Inverse

The difference quotient of the inverse of a function is found as the reciprocal of the its difference quotient.

In other words, for any function f defined at two adjacent nodes x and $x+\Delta x$ of a partition with $f(x)\neq f(x+\Delta x)$ so that its inverse function f^{-1} is defined at the two adjacent nodes $f(x)$ and $f(x+\Delta x)$ of a partition, we have the difference quotients (defined at the secondary nodes c and q within these edges of the two partitions respectively) satisfy:

$$\boxed{\frac{\Delta f^{-1}}{\Delta y}(q) = \frac{1}{\frac{\Delta f}{\Delta x}(c)}}$$

For the derivatives, we just take the limit $\Delta x\to 0$:

Theorem 4.9.4: Derivative of Inverse

The derivative of the inverse of a function is found as the reciprocal of the its derivative.

In other words, for any one-to-one function $y=f(x)$ differentiable at $x=a$, its

inverse $x = f^{-1}(y)$ is differentiable at $b = f(a)$, and we have:

$$\boxed{\frac{df^{-1}}{dy}(b) = \frac{1}{\frac{df}{dx}(a)}}$$

Warning!

The variables in the formula don't match.

Proof.

The formula follows from these limits:

$$\begin{array}{ccccl} \dfrac{\Delta x}{\Delta y} & \cdot & \dfrac{\Delta y}{\Delta x} & = 1 & \\ \downarrow & & \downarrow & & \\ \dfrac{dx}{dy} & \cdot & \dfrac{dy}{dx} & = 1 & \text{as } \Delta x \to 0,\ \Delta y \to 0\,. \end{array}$$

The fact that

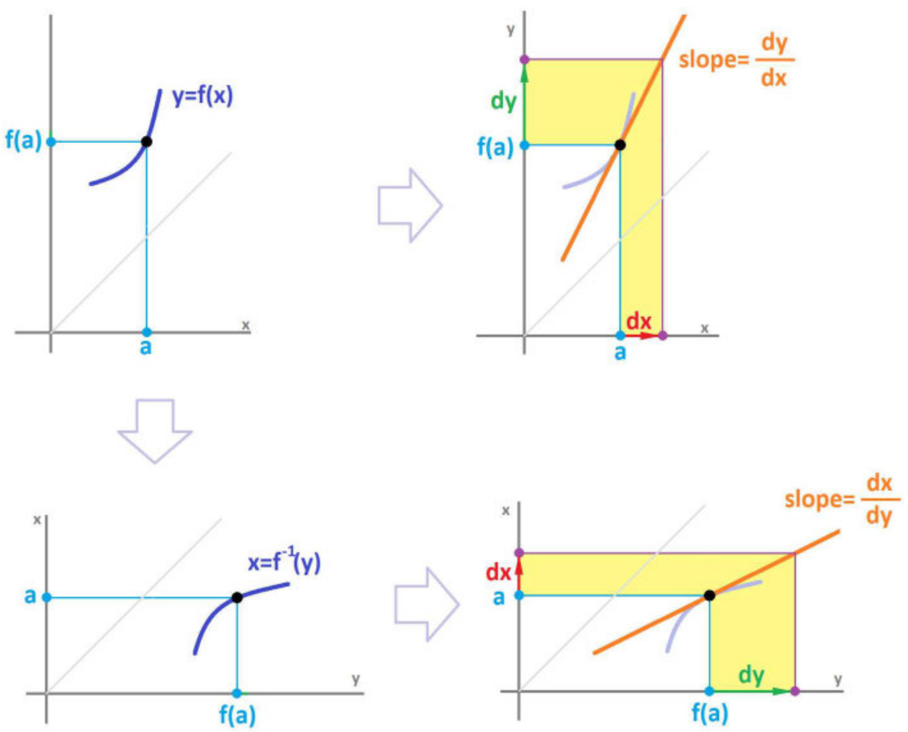
$$\Delta x \to 0 \implies \Delta y \to 0$$

follows from the continuity of f.

To be precise, we need to concentrate on a single point (a, b), where $b = f(a)$, on the graph of $y = f(x)$ and its tangent line. Then the derivatives

$$\left.\frac{dy}{dx}\right|_{x=a} \quad\text{and}\quad \left.\frac{dx}{dy}\right|_{y=b}$$

are indeed fractions and the reciprocals of each other:



Example 4.9.5: $\ln x$

Suppose we want to find the derivative of the logarithm. We'll use only its definition via the exponential function, as follows. We differentiate this equation, which amounts to the definition of the logarithm, of functions:

$$e^{\ln x} = x\,.$$

The flow chart below shows the dependencies:

$$x \ \to\ u = \ln x \ \to\ y = e^u$$

By the Chain Rule, we have:

$$\begin{array}{lll} (e^{\ln x})' & = (\ln x)' \cdot e^u & \\ & = (\ln x)' e^{\ln x} & \text{Now back-substitution.} \\ & = (\ln x)' x & \text{Cancellation.} \\ & = (x)' = 1\,. & \end{array}$$

Therefore,

$$(\ln x)' = \frac{1}{x}\,,$$

whenever $x > 0$.

We have a useful formula:

$$\boxed{(\ln x)' = \frac{1}{x}}$$

Similarly, we can find the derivatives of $\sin^{-1} x$, $\cos^{-1} x$, etc.

Example 4.9.6: another proof

For an alternative proof of the theorem, we differentiate the equation:

$$f^{-1}(f(x)) = x\,.$$

Then, by the *Chain Rule*, we have:

$$\frac{\Delta f^{-1}}{\Delta y}\frac{\Delta f}{\Delta x} = 1 \ \text{ and } \ \frac{df^{-1}}{dy}\frac{df}{dx} = 1\,.$$

The other equation produces the same result!

To make the formula in the theorem more useful, we add back-substitution:

Corollary 4.9.7: Derivative of Inverse With Back-Substitution

For any one-to-one function $y = f(x)$ differentiable at $x = a$, its inverse $x = f^{-1}(y)$ is differentiable at $b = f(a)$, and we have:

$$\boxed{\frac{df^{-1}}{dy}(b) = \frac{1}{\frac{df}{dx}(f^{-1}(b))}}$$

The formulas rewritten in the Lagrange notation are as follows:

$$\big(f^{-1}(f(a))\big)' = \frac{1}{f'(a)}\,.$$

A better version is below:

$$\boxed{\left(f^{-1}(b)\right)' = \frac{1}{f'(f^{-1}(b))}}$$

Example 4.9.8: $\arcsin x$

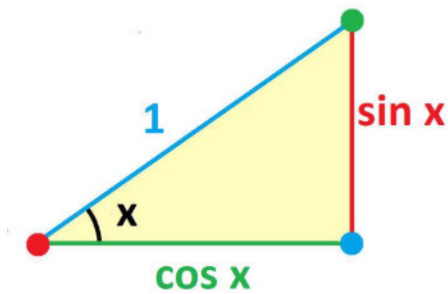
Find $(\sin^{-1} y)'$. There is no formula for this function, but its meaning is (for $-\pi/2 \le x \le \pi/2$) as follows:

$$y = \sin x, \text{ or } x = \sin^{-1} y\,.$$

Since $(\sin x)' = \cos x$, we conclude:

$$(\sin^{-1} y)' = \frac{1}{\cos x} = \frac{1}{\cos(\sin^{-1} y)}\,.$$

That *could* serve as the answer, but it's too cumbersome and should be simplified. We need to express $\cos x$ in terms of $\sin x$, which is y:



From the *Pythagorean Theorem*,

$$\sin^2 x + \cos^2 x = 1\,,$$

we conclude:

$$\begin{aligned}\cos x &= \sqrt{1 - \sin^2 x}\\ &= \sqrt{1 - y^2}\,.\end{aligned}$$

Therefore,

$$(\sin^{-1} y)' = \frac{1}{\sqrt{1 - y^2}}\,.$$

We can apply the theorem to other trigonometric functions. These are the results:

$$\boxed{\begin{aligned}(\sin^{-1} x)' &= \frac{1}{\sqrt{1-x^2}}\\ (\cos^{-1} x)' &= -\frac{1}{\sqrt{1-x^2}}\\ (\tan^{-1} x)' &= \frac{1}{1+x^2}\end{aligned}}$$

Exercise 4.9.9

Prove the formulas.

Exercise 4.9.10

Since $(\sin^{-1} x)' = -(\cos^{-1} x)'$, does it mean that $\sin^{-1} x = -\cos^{-1} x$?

Even though it is impossible to see it in these formulas, we repeat:

▶ The derivatives of inverses are the reciprocals of each other.

We can rewrite the Inverse Rule in the Leibniz notation in terms of the variables only:

$$\frac{dx}{dy} = \frac{1}{\frac{dy}{dx}}\,.$$

An even better version is below:

$$\boxed{\frac{dx}{dy} \cdot \frac{dy}{dx} = 1}$$

The *Power Formula* is a result from last chapter with an incomplete (integers only) proof. We start with reciprocal powers because they are roots, i.e., the inverses of the powers:

Theorem 4.9.11: Derivative of Reciprocal Powers

For any positive integer n, we have:

$$\boxed{\frac{dy^{\frac{1}{n}}}{dy} = \frac{1}{n} y^{\frac{1}{n}-1}}$$

Proof.

The inverse of $x = y^{\frac{1}{n}}$ is $y = x^n$. Therefore,

$$\frac{dy^{\frac{1}{n}}}{dy} = \frac{1}{\frac{dx^n}{dx}} = \frac{1}{nx^{n-1}} = \frac{1}{n\left(y^{\frac{1}{n}}\right)^{n-1}} = \frac{1}{ny^{\frac{n-1}{n}}} = \frac{1}{n} y^{\frac{1}{n}-1}\,.$$

We use the above formula to prove the general one:

Theorem 4.9.12: Derivative of Rational Powers

For any positive integers n and m, we have:

$$\boxed{\frac{dy^{\frac{m}{n}}}{dy} = \frac{m}{n} y^{\frac{m}{n}-1}}$$

Exercise 4.9.13

Prove the theorem.

Warning!

The fact that $(x^\pi)' = \pi x^{\pi-1}$ will remain unproven.

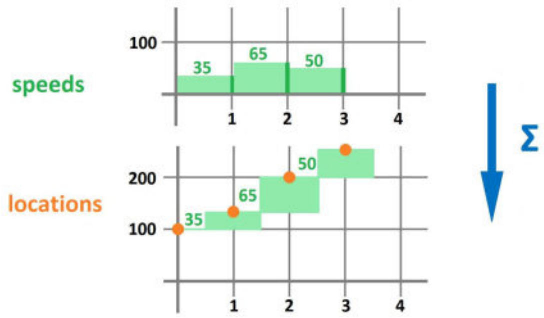
4.10. *Reversing differentiation*

We have encountered the following question several times by now:

▶ When we know the velocity at every moment of time, how do we find the location?

The question applies equally to the velocity acquired from the location as its difference quotient or as its derivative. To answer the question, we need to "reverse-engineer" their effect on a function.

Another simple example that we have seen several times is that of a broken speedometer and its "inverse", the problem of a broken odometer:



It is solved by simple addition.

The problem is simpler: If we know the displacements during each of the time periods, can we find our location? Just add them together to find the total displacement! This is about the *difference*.

Suppose we have a function $y = g(x)$ defined at the secondary nodes, c, of a partition. How do we find a function $y = f(x)$ defined at the nodes, x, of the partition so that g is its difference:

$$\Delta f(c) = g(c)?$$

In other words, this is what we face:

$$\boxed{\text{Solve: } \Delta f = g}$$

Solving this equation isn't hard.

Suppose this function g is known but f isn't, except for one (initial) value: $y_0 = f(a)$. Then the above equation becomes:

$$\Delta f(c_1) = f(x_0 + \Delta x_1) - f(x_0) = g(c_1)\,,$$

and we can solve it:

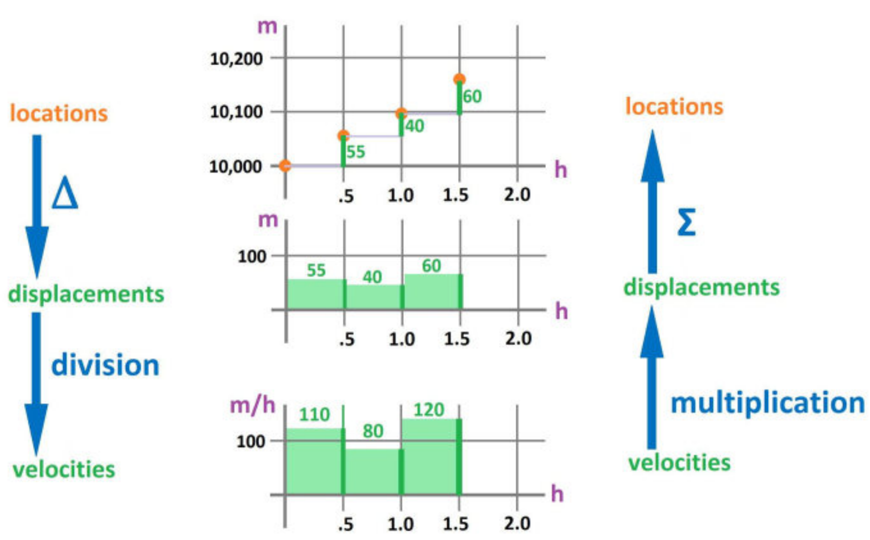
$$f(x_1) = f(x_0 + \Delta x_1) = f(x_0) + g(c_1)\,.$$

We continue in this manner and find the rest of the values of f:

$$f(x_{k+1}) = f(x_k + \Delta x_k) = f(x_k) + g(c_k)\,.$$

This formula is *recursive*: We need to know the last value of f in order to find the next. Though not an explicit formula, the solution is very simple!

Now, the *difference quotient*.



Suppose we have a function $y = v(x)$ defined at the secondary nodes, c, of a partition. How do we find a function $y = p(x)$ defined at the nodes, x, of the partition so that v is its difference quotient:

$$\frac{\Delta p}{\Delta x}(c) = v(c)?$$

In other words, this is what we face:

$$\boxed{\text{Solve: } \frac{\Delta p}{\Delta x} = v}$$

We just follow exactly the process above.

Suppose this function v is known but p isn't, except for one (initial) value: $y_0 = p(a)$. Then the above equation becomes:

$$\frac{\Delta p}{\Delta x}(c_1) = \frac{p(x_0 + \Delta x_1) - p(x_0)}{\Delta x_1} = v(c_1)\,,$$

and we can solve it:

$$p(x_1) = p(x_0 + \Delta x_1) = p(x_0) + v(c_1)\Delta x_1\,.$$

We continue in this manner and find the rest of the values of f:

$$p(x_{k+1}) = p(x_k + \Delta x_k) = p(x_k) + v(c_k)\Delta x_k\,.$$

This formula is also recursive, but, within this limitation, the problem of reversing the effect of the difference quotient is solved!

Just as in the last chapter, we initially face the problem:

$$\text{position} \ \to \ \text{velocity} \ \to \ \text{acceleration}$$

We now go in reverse:

$$\text{position} \ \leftarrow \ \text{velocity} \ \leftarrow \ \text{acceleration}$$

These two problems are similar to the one of finding the *inverse* of a function. This is how inverse functions appear in algebra; they come from solving equations, for x:

$$\begin{array}{lllll} x^2 & = 4 & \Longrightarrow & x = 2 & \text{via } \sqrt{\cdot} \\ 2^x & = 8 & \Longrightarrow & x = 3 & \text{via } \log_2(\cdot) \\ \sin x & = 0 & \Longrightarrow & x = 0 & \text{via } \sin^{-1}(\cdot) \end{array}$$

In other words, what do we do if we know the output of a function and want to know the input? Initially, we can only *recollect* a past experience with a function! For a repeated use, we develop the *inverse* of the function.

Similarly, what do we do if we know the result of differentiation and want to know where it came from? We need to solve these, for example:

$$\begin{array}{llll} f' & = 2x & \Longrightarrow & f & = x^2 \\ f' & = \cos x & \Longrightarrow & f & = \sin x \\ f' & = e^x & \Longrightarrow & f & = e^x \end{array}$$

In other words, this is what we face:

$$\boxed{\text{Solve: } \frac{dp}{dx} = v}$$

This equation isn't as easy as the last.

Initially, we can only *recollect* a past experience with differentiation! For a repeated use, we will need to develop the *inverse* of the function (Chapter 5). Because the derivative is not a fraction but a limit of a fraction, there is no single, even recursive, formula.

This process is called *anti-differentiation*.

Example 4.10.1: free fall

The importance of this "inverse" problem stems from the need to find location from velocity or velocity from acceleration.

For example, this is what we derive from our experience with differentiation. To understand free fall, we will need to prove the following two statements:

1. For the horizontal component: The acceleration is zero. $\Longrightarrow$ The velocity is constant. $\Longrightarrow$ The location is a linear function.
2. For the vertical component: The acceleration is constant. $\Longrightarrow$ The velocity is a linear function. $\Longrightarrow$ The location is a quadratic function.

We illustrate the idea of anti-differentiation with a diagram:

$$\begin{array}{ccccc} x^2 & \to & \boxed{\dfrac{d}{dx}} & \to & 2x \\ 2x & \to & \boxed{\left(\dfrac{d}{dx}\right)^{-1}} & \to & x^2 \end{array}$$

We've found one solution. Are there any others? Yes, $(x^2+1)' = 2x$. And more:

$$\begin{array}{ccc} & & x^2+1 \\ & \nearrow & \\ 2x & \to & x^2 \\ & \searrow & \\ & & x^2-1 \end{array}$$

As a function – a function of functions – $\dfrac{d}{dx}$ isn't one-to-one!

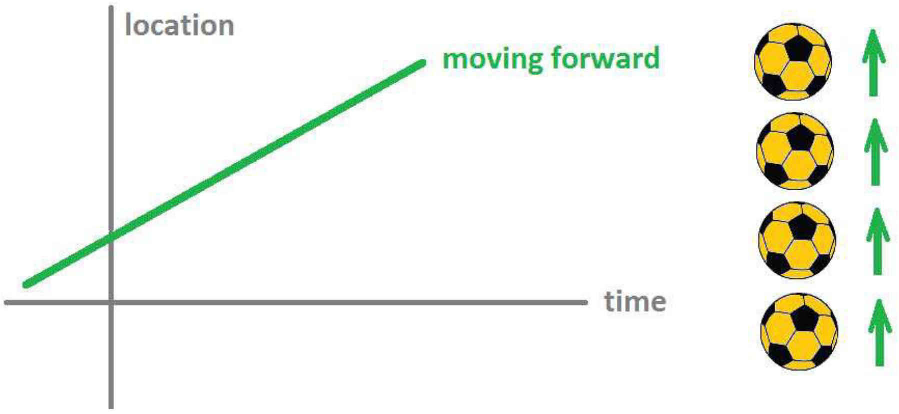
Exercise 4.10.2

We can make any function one-to-one by restricting its domain. How would that work for differentiation?

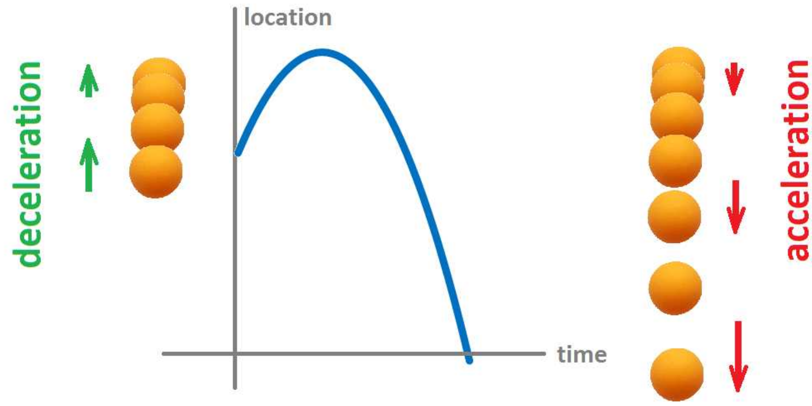
We apply these ideas to a specific problem about motion.

4.11. *Shooting a cannon*

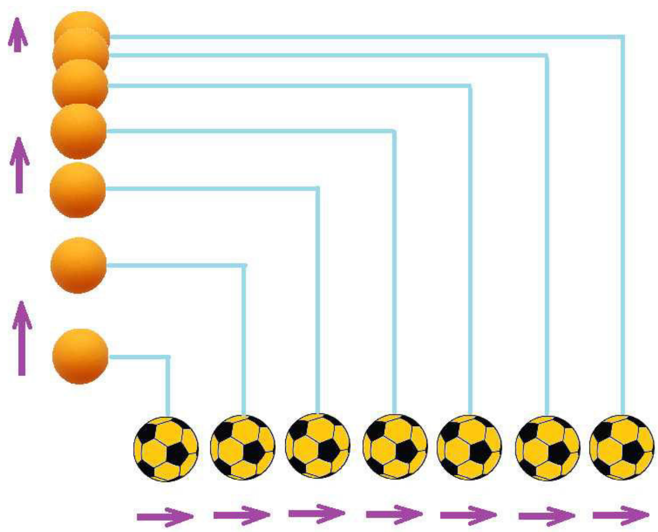
A soccer ball rolling on a horizontal plane will have a constant velocity:



A ping-pong ball thrown up in the air goes up, slows down until it stops for an instant, and then accelerates toward the surface:

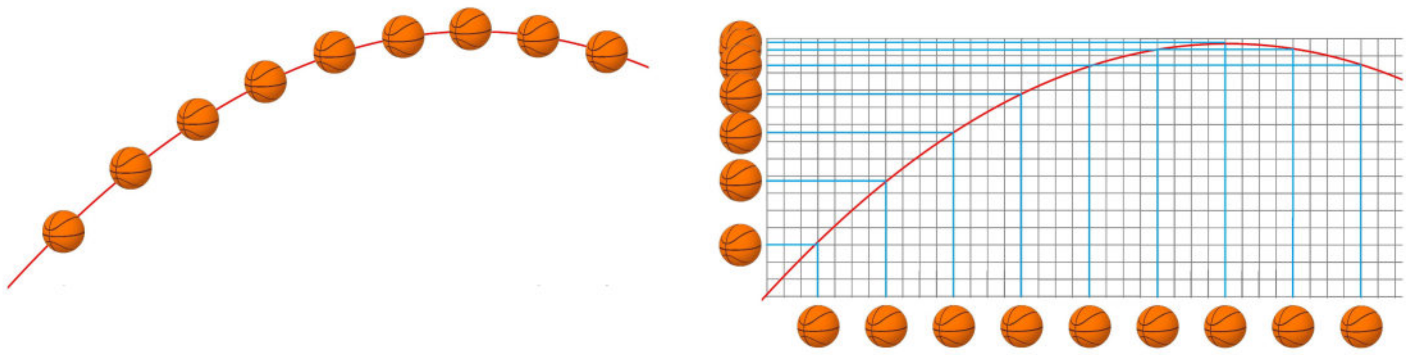


What if we do both: We roll a soccer ball horizontally and throw a ping-pong ball vertically? Let's try to follow both balls at the same time:



We'd have to fly through the air as if thrown at an angle!

Our understanding is that a thrown ball moves in both vertical and horizontal directions, simultaneously and independently:



The dynamics is very different:

1. In the horizontal direction, as there is no force changing the velocity, the latter remains constant.
2. Meanwhile, the vertical velocity is constantly changed by the force of gravity.

Let's now use these descriptions to represent the motion mathematically.

Recall how earlier in this chapter we used these *difference quotients* to find velocity and then the acceleration from the location:

$$v_n = \frac{\Delta p}{\Delta t} = \frac{p_{n+1} - p_n}{h} \quad \text{and} \quad a_n = \frac{\Delta v}{\Delta t} = \frac{v_{n+1} - v_n}{h},$$

where h is the increment of time.

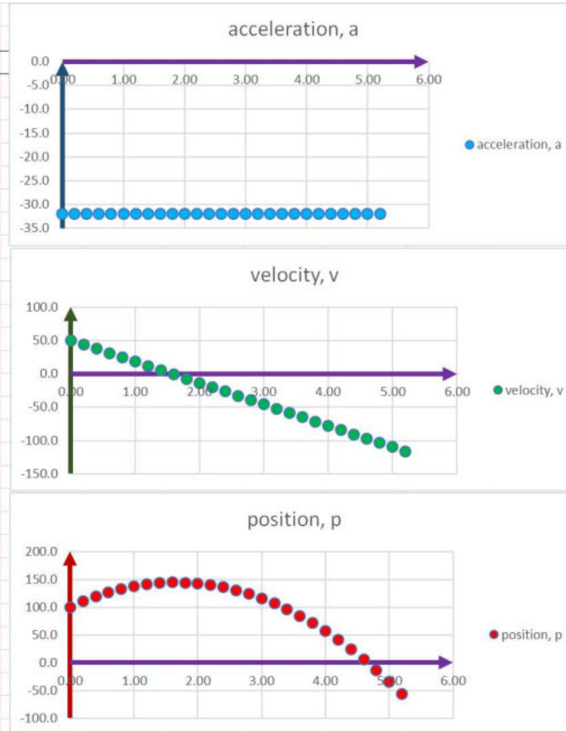
These formulas have also been solved for p_{n+1} and v_{n+1} respectively in order to be able to model location as a function of time:

$$v_{n+1} = v_n + ha_n \quad \text{and} \quad p_{n+1} = p_n + hv_n \,.$$

These recursive formulas are called the *Riemann sums*.

This is what the results might look like:

h	t	a	v	p
0.20				
intial:	0.00	-32.00	50.00	100.00
	0.20	-32.00	43.60	110.00
	0.40	-32.00	37.20	118.72
	0.60	-32.00	30.80	126.16
	0.80	-32.00	24.40	132.32
	1.00	-32.00	18.00	137.20
	1.20	-32.00	11.60	140.80
	1.40	-32.00	5.20	143.12
	1.60	-32.00	-1.20	144.16
	1.80	-32.00	-7.60	143.92
	2.00	-32.00	-14.00	142.40
	2.20	-32.00	-20.40	139.60
	2.40	-32.00	-26.80	135.52
	2.60	-32.00	-33.20	130.16
	2.80	-32.00	-39.60	123.52
	3.00	-32.00	-46.00	115.60
	3.20	-32.00	-52.40	106.40
	3.40	-32.00	-58.80	95.92
	3.60	-32.00	-65.20	84.16
	3.80	-32.00	-71.60	71.12
	4.00	-32.00	-78.00	56.80
	4.20	-32.00	-84.40	41.20
	4.40	-32.00	-90.80	24.32
	4.60	-32.00	-97.20	6.16
	4.80	-32.00	-103.60	-13.28
	5.00	-32.00	-110.00	-34.00
	5.20	-32.00	-116.40	-56.00
	5.40	-32.00	-122.80	-79.28
	5.60	-32.00	-129.20	-103.84
	5.80	-32.00	-135.60	-129.68



This time we have *two* such sequences, one for horizontal and one for vertical.

We construct the Cartesian coordinate system in the most convenient way:

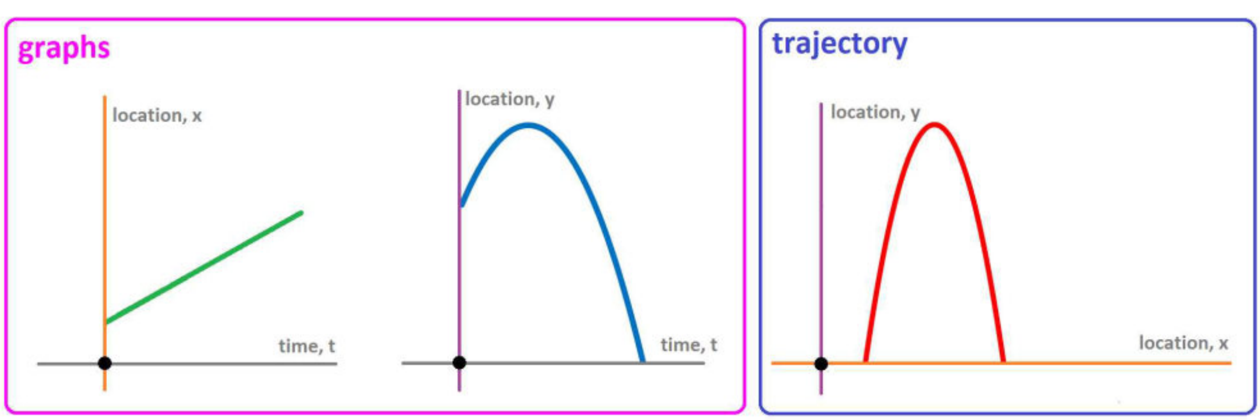
- The x-axis is horizontal.
- The y-axis is vertical.

However, we abandon the familiar $y = f(x)$ setup! We have *three variables* now:

- t is time.
- x is the horizontal dimension, the depth.
- y is the vertical dimension, the height.

Either of the two *spatial* variables depends on the *temporal* variable.

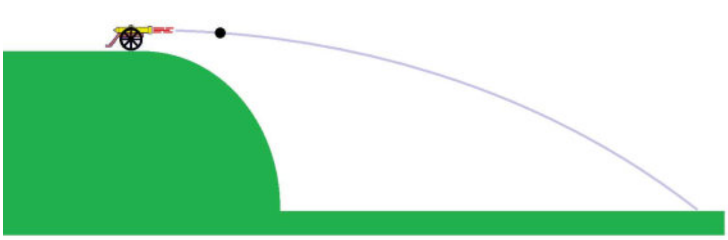
Their *graphs* are plotted below (left):

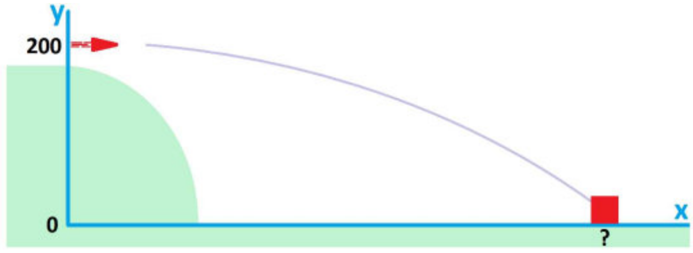


Meanwhile, the *path* of the ball will appear to an observer as a curve in the, vertically aligned, xy-plane (right).

Historically, one of the very first applications of calculus was in *ballistics*. Before calculus, one had to resort to trial and error and watching where the cannonballs were landing. A well-designed test may provide one with a table (i.e., a function) that gives the shot length for each angle of the barrel. However, such a reference table may prove useless when one is to shoot from an elevated position, or at an elevated target, or over an obstacle.

> ▶ *PROBLEM:* From a 200-foot elevation, a cannon is fired horizontally at 200 feet per second. How far will the cannonball go?





We will find the whole path!

Let $h = \Delta t$ be the increment of time. We have these *six* sequences with the difference quotients computed four times:

	horizontal	vertical	
position	x_n	y_n	
velocity	$v_n = \dfrac{x_{n+1} - x_n}{h}$	$u_n = \dfrac{y_{n+1} - y_n}{h}$	DQ
acceleration	$a_n = \dfrac{v_{n+1} - v_n}{h}$	$b_n = \dfrac{u_{n+1} - u_n}{h}$	DQ

Now, from the point of modeling, the derivation should go in the opposite direction. We go in reverse: the velocity and then the location from the acceleration. When we solve the above equations, we end up with these *four* recursive formulas (the Riemann sums) for our six sequences:

	horizontal	vertical	
acceleration	a_n	b_n	
velocity	$v_{n+1} = v_n + ha_n$	$u_{n+1} = u_n + hb_n$	RS
position	$x_{n+1} = x_n + hv_n$	$y_{n+1} = y_n + hu_n$	RS

Now in the specific case of *free fall*, there is just one force, the gravity. Therefore, the horizontal acceleration is zero and the vertical acceleration is constant (feet per second squared):

$$a = 0, \ b = -32 .$$

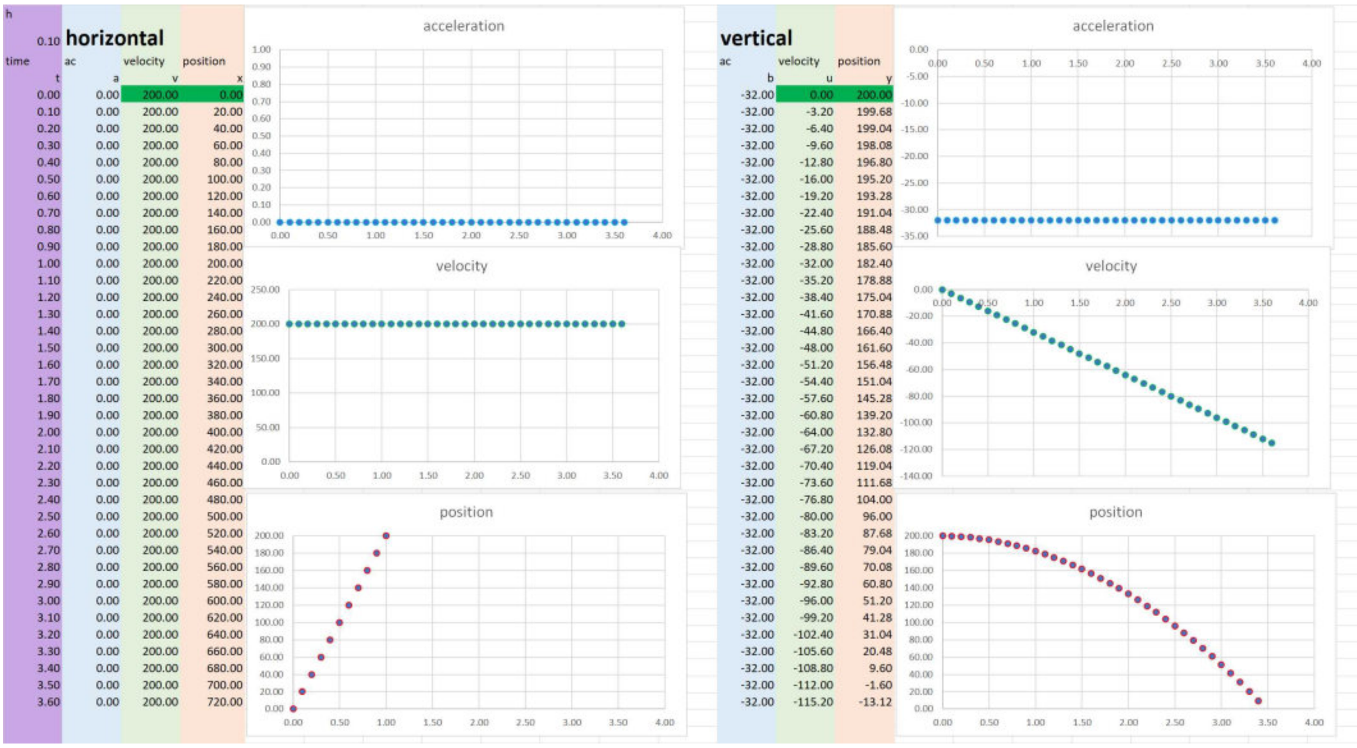
Next, we acquire the initial conditions:

	x	y
initial location:	$x_0 = 0$	$y_0 = 200$
initial velocity:	$v_0 = 200$	$u_0 = 0$

These four numbers serve as the *initial terms* of our four sequences:

time		horizontal	vertical
t_0	acceleration	$a_0 = 0$	$b_0 = -32$
	velocity	$v_1 = 200 + .1 \cdot 0$	$u_1 = 0 + .1 \cdot (-32)$
	position	$x_1 = 0 + .1 \cdot 200$	$y_1 = 200 + .1 \cdot 0$
$t_1 = t_0 + h$	acceleration	$a_1 = 0$	$b_1 = -32$
	velocity	$v_2 = v_1 + ha_1$	$u_2 = u_1 + hb_1$
	position	$x_2 = x_1 + hv_1$	$y_3 = y_1 n + hu_1$
$t_2 = t_1 + h$	...	...	

We use the formulas to evaluate the location every $h = .1$ second. We take the spreadsheet presented above, copy the columns for acceleration, velocity, and position, and paste next:



time t	ac a	velocity v	position x	ac b	velocity u	position y
0.00	0.00	200.00	0.00	-32.00	0.00	200.00
0.10	0.00	200.00	20.00	-32.00	-3.20	199.68
0.20	0.00	200.00	40.00	-32.00	-6.40	199.04
0.30	0.00	200.00	60.00	-32.00	-9.60	198.08
0.40	0.00	200.00	80.00	-32.00	-12.80	196.80
0.50	0.00	200.00	100.00	-32.00	-16.00	195.20
0.60	0.00	200.00	120.00	-32.00	-19.20	193.28
0.70	0.00	200.00	140.00	-32.00	-22.40	191.04
0.80	0.00	200.00	160.00	-32.00	-25.60	188.48
0.90	0.00	200.00	180.00	-32.00	-28.80	185.60
1.00	0.00	200.00	200.00	-32.00	-32.00	182.40
1.10	0.00	200.00	220.00	-32.00	-35.20	178.88
1.20	0.00	200.00	240.00	-32.00	-38.40	175.04
1.30	0.00	200.00	260.00	-32.00	-41.60	170.88
1.40	0.00	200.00	280.00	-32.00	-44.80	166.40
1.50	0.00	200.00	300.00	-32.00	-48.00	161.60
1.60	0.00	200.00	320.00	-32.00	-51.20	156.48
1.70	0.00	200.00	340.00	-32.00	-54.40	151.04
1.80	0.00	200.00	360.00	-32.00	-57.60	145.28
1.90	0.00	200.00	380.00	-32.00	-60.80	139.20
2.00	0.00	200.00	400.00	-32.00	-64.00	132.80
2.10	0.00	200.00	420.00	-32.00	-67.20	126.08
2.20	0.00	200.00	440.00	-32.00	-70.40	119.04
2.30	0.00	200.00	460.00	-32.00	-73.60	111.68
2.40	0.00	200.00	480.00	-32.00	-76.80	104.00
2.50	0.00	200.00	500.00	-32.00	-80.00	96.00
2.60	0.00	200.00	520.00	-32.00	-83.20	87.68
2.70	0.00	200.00	540.00	-32.00	-86.40	79.04
2.80	0.00	200.00	560.00	-32.00	-89.60	70.08
2.90	0.00	200.00	580.00	-32.00	-92.80	60.80
3.00	0.00	200.00	600.00	-32.00	-96.00	51.20
3.10	0.00	200.00	620.00	-32.00	-99.20	41.28
3.20	0.00	200.00	640.00	-32.00	-102.40	31.04
3.30	0.00	200.00	660.00	-32.00	-105.60	20.48
3.40	0.00	200.00	680.00	-32.00	-108.80	9.60
3.50	0.00	200.00	700.00	-32.00	-112.00	-1.60
3.60	0.00	200.00	720.00	-32.00	-115.20	-13.12

Of course, for the horizontal values, we replace acceleration with $a = 0$.

Example 4.11.1: how far

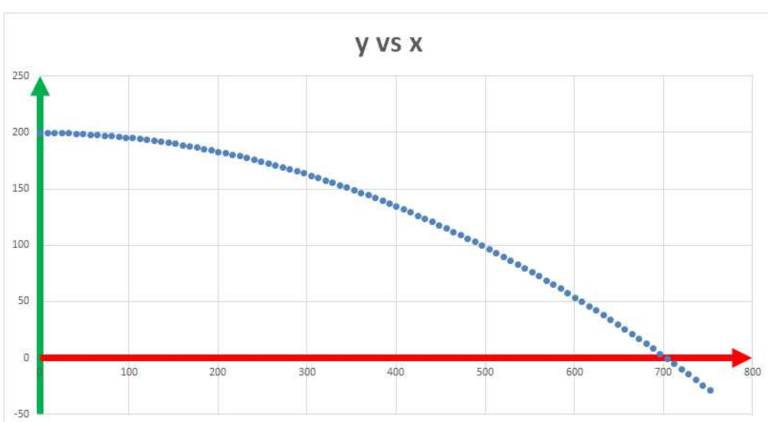
To find when and where the ball hits the ground, we scroll down to find the row with y close to 0.

3.20	0.00	200.00	640.00	-32.00	-102.40	31.04
3.30	0.00	200.00	660.00	-32.00	-105.60	20.48
3.40	0.00	200.00	680.00	-32.00	-108.80	9.60
3.50	0.00	200.00	700.00	-32.00	-112.00	-1.60
3.60	0.00	200.00	720.00	-32.00	-115.20	-13.12

It happens about $t = 3.5$. Then, the value of x at the time is about $x = 700$. How fast is it going at the time? We use the velocity values from the same row:

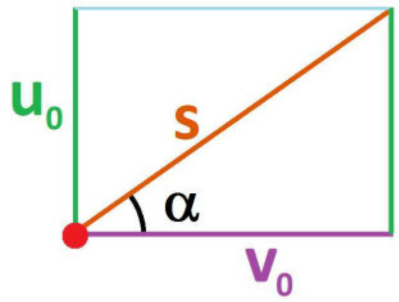
$$\sqrt{200^2 + 108^2} \approx 227.3 \quad \text{feet per second.}$$

Now we combine the x- and y-columns to plot the path:



With the spreadsheet, we can ask and answer a variety of questions about such motion. But first, let's introduce the *angle* of the barrel of the cannon into the model.

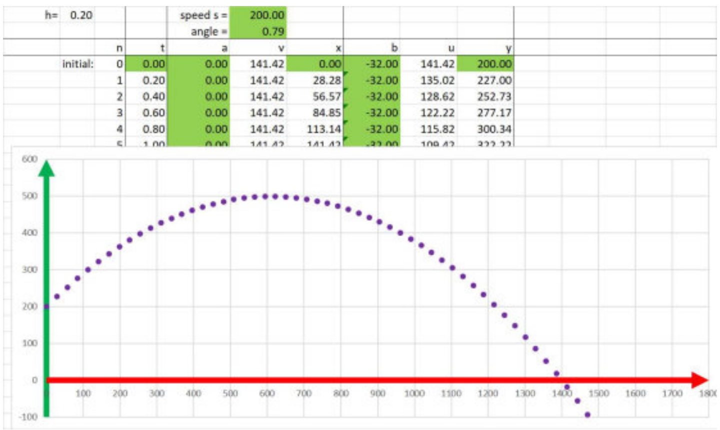
The velocity of 200 feet per second we have been using is the "muzzle velocity", i.e., the *speed*, s, with which the cannonball leaves the muzzle – no matter what the angle, α, is. That's where the *initial* horizontal and the vertical velocities come from:



The formulas come from trigonometry:

$$v_0 = s\cos\alpha \quad \text{and} \quad u_0 = s\sin\alpha\,.$$

We use them below to provide the initial values of the two velocities:



h= 0.20			speed s =	200.00				
			angle =	0.79				
	n	t	a	v	x	b	u	y
initial:	0	0.00	0.00	141.42	0.00	-32.00	141.42	200.00
	1	0.20	0.00	141.42	28.28	-32.00	135.02	227.00
	2	0.40	0.00	141.42	56.57	-32.00	128.62	252.73
	3	0.60	0.00	141.42	84.85	-32.00	122.22	277.17
	4	0.80	0.00	141.42	113.14	-32.00	115.82	300.34

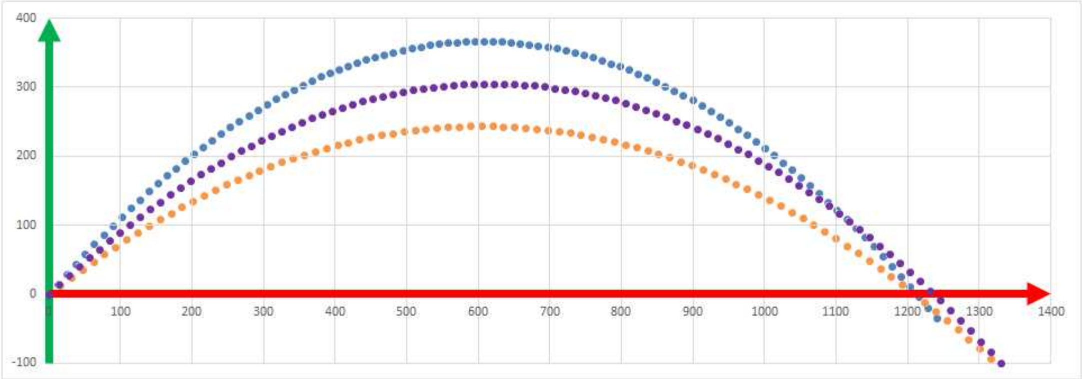
We can freely enter the data for the following (highlighted in green):

- the initial speed
- the initial angle
- the initial location
- all accelerations

The rest is computed according to the same formulas as before.

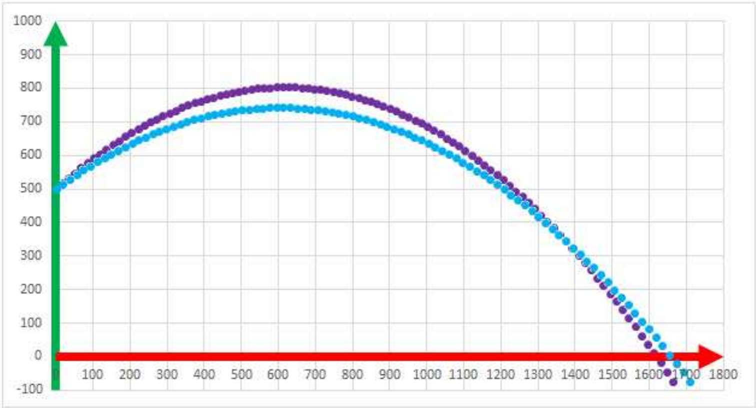
Example 4.11.2: longest shot

Is it really true that 45 degrees is the best angle to shoot for a longer distance? To test the idea, we try to shoot with an angle just above and just below:



It appears that the one in the middle is the best, but we can't prove this with just the numerical methods.

Now, what if we try to shoot from a hill again, say, 500 feet high?



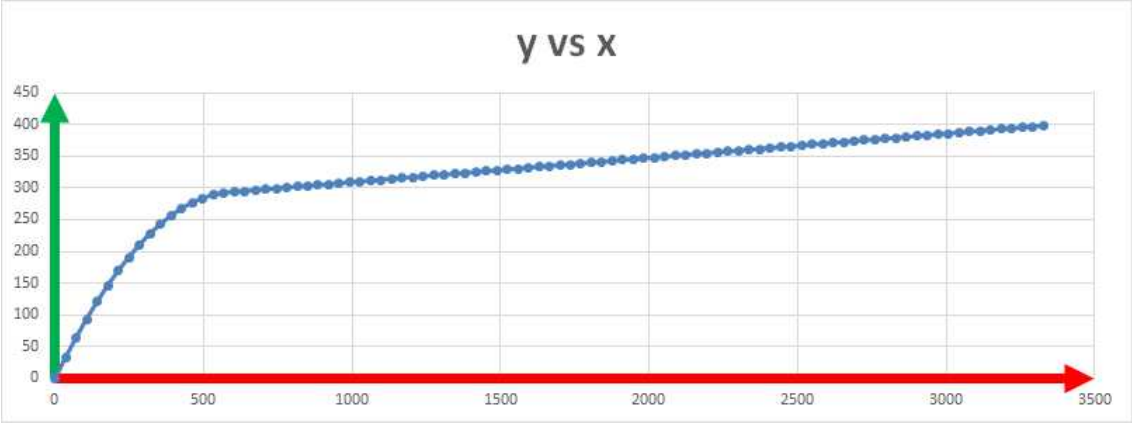
It's not the best anymore!

Exercise 4.11.3

Show that the best shot will become more and more flat as the elevation grows.

Example 4.11.4: variable gravity

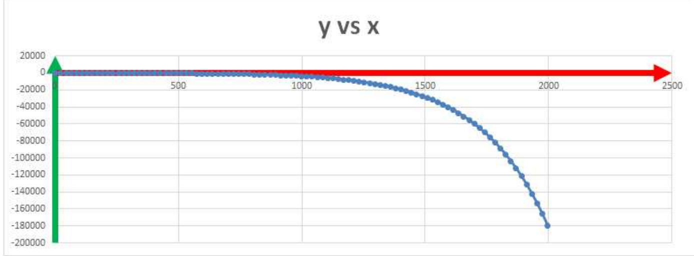
What happens if the gravity suddenly disappears? In the column for the vertical acceleration, we just replace -32 with 0 after a few rows:



The cannonball flies off on a *tangent*.

Example 4.11.5: variable gravity

What happens if the gravity starts to increase? We try to increase the down acceleration 1 foot per second squared per second:



The trajectory looks steeper and steeper, but is there a *vertical asymptote*? We can't answer with just the numerical methods.

Example 4.11.6: horizontal gravity

What happens if the gravity is horizontal instead? The motion will be along a parabola that lies on its side, of course. But what if there is both vertical and horizontal gravity? Let's modify the acceleration columns accordingly by replacing 0's with -32 in the horizontal acceleration column:

n	t	Ax	Vx	x	Ay	Vy	y
0	0.00	32.00	-100.00	0.00	-32.00	173.21	200.00
1	0.10	32.00	-96.80	-9.68	-32.00	170.01	217.00
2	0.20	32.00	-93.60	-19.04	-32.00	166.81	233.68

Is this a parabola?

Exercise 4.11.7

Explain the results in the last example.

In spite of these numerous examples, we can only do one at a time! The conclusions we draw are also specific to these situations (accelerations, initial conditions, etc.).

This is why we now consider the *continuous* case, i.e., we take the limit of everything above:

$$\boxed{h = \Delta t \to 0}$$

This time, instead of six *sequences*, we have these six *functions* of time:

x, the depth, the horizontal location	$v = x'$, the horizontal velocity	$a = v'$, the horizontal acceleration
y, the height, the vertical location	$u = y'$, the vertical velocity	$b = u'$, the vertical acceleration

There is no time increment as a parameter anymore!

Now the specific case of free fall:

$$a = 0,\ b = -g\,.$$

We have learned in this chapter that:

1. The derivative of a quadratic polynomial is linear.
2. The derivative of a linear polynomial is constant.

We will show in Chapter 5 that, conversely, we have:

1. The only function the derivative of which is linear is a quadratic polynomial.
2. The only function the derivative of which is constant is a linear polynomial.

From the latter, we conclude about *free fall*:

1. The horizontal position $x = x(t)$ is linear.
2. The vertical position $y = y(t)$ is quadratic.

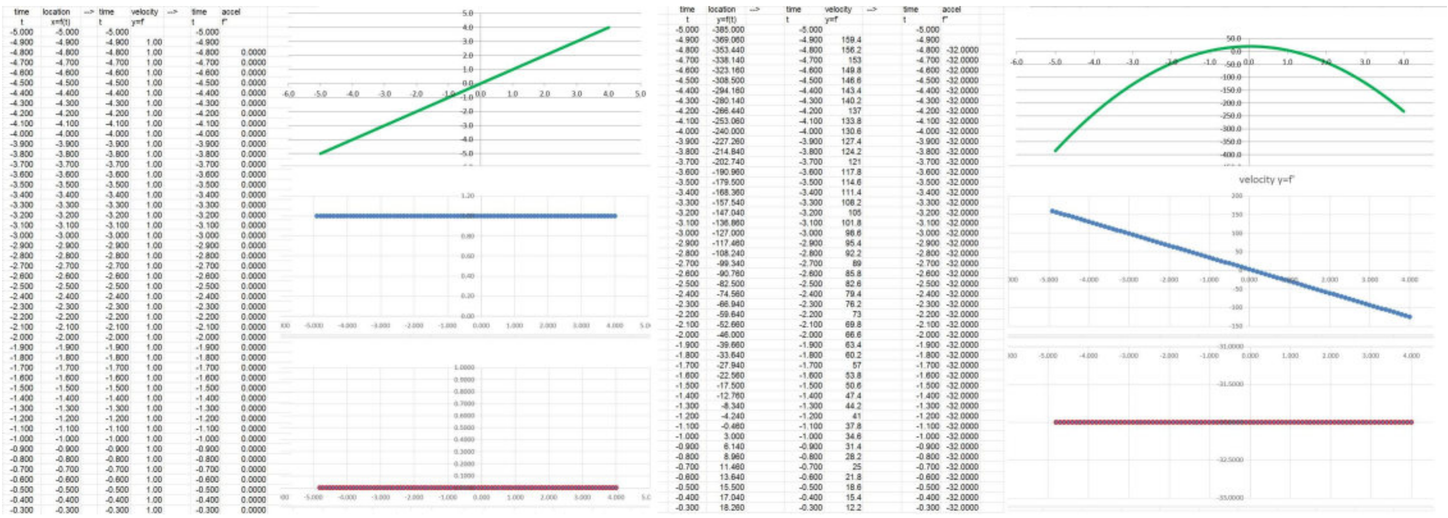
What makes these two functions specific are the *initial conditions*:

x_0, the initial depth, $x_0 = x(0)$	v_0, the initial horizontal component of velocity, $v(0) = \frac{dx}{dt}\Big\|_{t=0}$
y_0, the initial height, $y_0 = y(0)$	u_0, the initial vertical component of velocity, $u(0) = \frac{dy}{dt}\Big\|_{t=0}$

Therefore, we have:

$$\boxed{\begin{array}{llll} x & = x_0 & +v_0 t & \\ y & = y_0 & +u_0 t & -\frac{1}{2}gt^2 \end{array}}$$

These two equations allow us to solve a variety of problems about motion. We carry this out for x and y separately and the results are shown in the spreadsheet:



Example 4.11.8: how far

Let's revisit the problem about a specific shot we solved numerically. Our equations become:

$$\begin{array}{lll} x & = & 200t \\ y & = 200 & \qquad -16t^2 \end{array}$$

Now, analytically, the height at the end is y_0, so to find *when* it happened, we set $y = 0$, or

$$200 - 16t^2 = 0\,,$$

and solve for t. Then, the time of landing is:

$$t_1 = \sqrt{\frac{200}{16}} = \frac{5\sqrt{2}}{2}\,.$$

To find *where* it happened, we substitute this value of t into x; the location is:

$$x_1 = 200t_1 = 200\frac{5\sqrt{2}}{2} \approx 707\,.$$

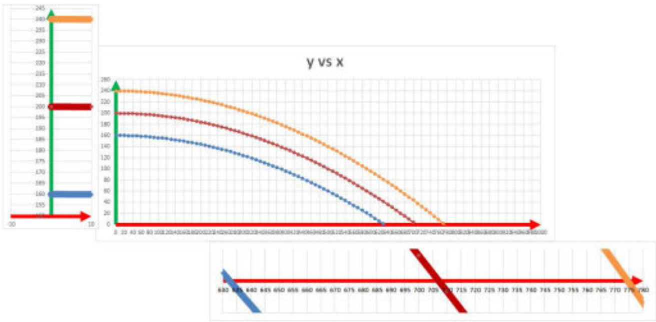
The result matches our estimate!

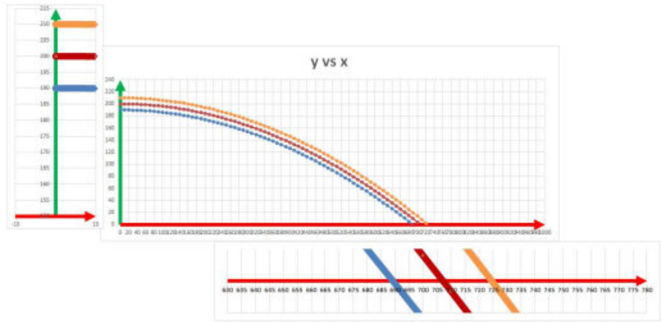
We will prove in Chapter 6 that the longest distance is achieved when shot at 45 degrees.

Example 4.11.9: accuracy

From the practical point of view, no shot is perfectly accurate. Even when the mathematics is seen as perfect, our limited knowledge of the many parameters that affect the accuracy of our shot will make us under- or overestimate the final location of the cannonball.

Let's take just one: the accuracy of our measurement of the height of the hill. If the height, H, varies, then so will the placement, D, of the shot. The good news is that the smaller error in the former will lead to a smaller error in the latter!





In fact, we can achieve *any* required degree of accuracy, ε, of our shot if we can ensure a sufficient accuracy, δ, of the measurement of the height of the hill. In other words, the dependence of D on H is *continuous*: The distance of the shot depends continuously on the initial elevation.

Exercise 4.11.10

Prove this statement algebraically.

Exercise 4.11.11

What if instead of target shooting this was a game of tennis?

Exercise 4.11.12

Prove that the distance of the shot depends continuously on the initial *angle*.

It so happens that the path is the graph of y as a function of x! In that case, this dependence can also be found by substitution of t as a function of x into $y(t)$.

Is the path a parabola? It looks like one when plotted point by point (as above), but let's do the algebra now. From the equation for x,

$$x = x_0 + v_0 t\,,$$

we conclude that

$$t = (x - x_0)/v_0$$

whenever $v_0 \neq 0$. Then, we substitute into the equation for y:

$$\boxed{y = y_0 + \frac{u_0}{v_0}(x - x_0) - \frac{g}{2v_0^2}(x - x_0)^2}$$

This is a quadratic function! We have proven it:

► *The trajectory of the cannonball is a parabola.*

That's why the path traced in the sky is curved this way.

Exercise 4.11.13

What is the path when $v_0 = 0$?

Exercise 4.11.14

Show that the coefficient $\frac{u}{v}$ is the tangent of the angle of the shot.

Chapter 5: The main theorems of differential calculus

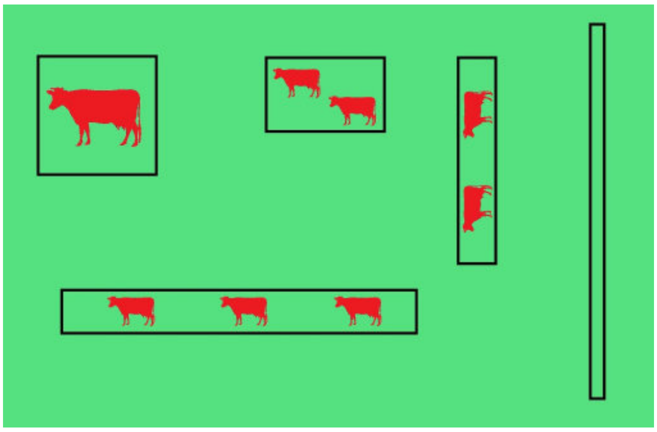
Contents

5.1. Monotonicity and extreme points

This is how we have been able to solve optimization problems so far.

Example 5.1.1: farmer

Let's review the following, possibly familiar, problem (seen in Volume 1, Chapter 1PC-2): A farmer with 100 yards of fencing material wants to build as large a rectangular enclosure as possible for his cattle:



We define the variables: the width W (the depth $D = 50 - W$) and the area A. The the area is a function of W:

$$A = W(50 - W)\,.$$

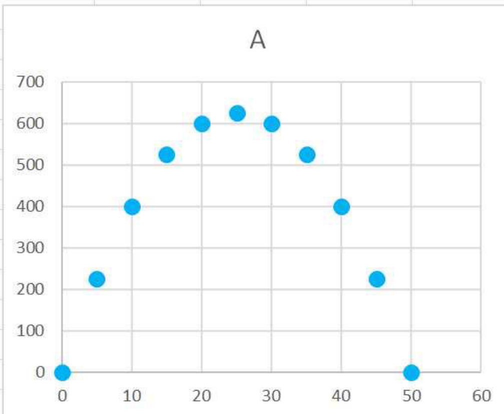
We choose a meaningful domain for the function: $[0, 50]$.

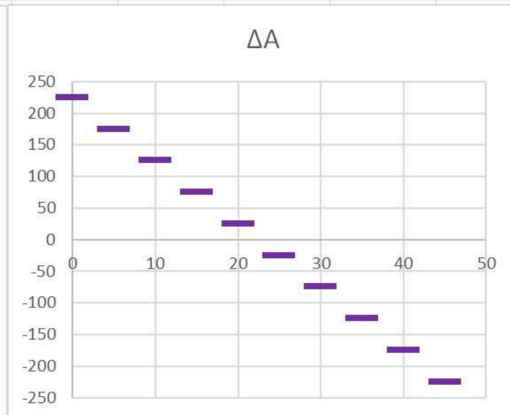
Then the problem becomes:

- What is the largest output of the function $A(W) = -W^2 + 50W$ with $0 \le W \le 50$?

In its simplest interpretation, the answer to the question is found by examining the data produced by this formula and choosing the largest value from the A-column (the outputs):

5			
n	W	A	ΔA
1	0	0	
2	5	225	225
3	10	400	175
4	15	525	125
5	20	600	75
6	25	625	25
7	30	600	-25
8	35	525	-75
9	40	400	-125
10	45	225	-175
11	50	0	-225





The maximal value of the area appears to be $25 \cdot 25 = 625$.

To confirm, we evaluate the monotonicity of the dependence of A on W. We examine the data produced by this formula and notice a pattern of growth and decline of the values in the A-column.

One can also watch the sign of the *differences* of A; they are easily computed. Then, the values are increasing when the difference is positive! These are the points where the *difference quotients* of our function are positive too. We observe that on the interval $[0, 25]$, the difference quotient is positive and elsewhere is negative.

Now, here is a slightly different approach. Examining the graph reveals that the maximum value lies somewhere in the area where the graph is the *flattest*. In other words, this is where the slopes of the secant lines are closest to zero. But these slopes are the difference quotients of our function. Let's find them. From the *Power Formula*, we have:

$$\frac{\Delta A}{\Delta W} = -2W - h + 50\,.$$

We conclude that on interval $[25, 25+h]$, the difference quotient is $-h$. This is potentially the number closest to 0 (provided h is small enough). We have arrived at the same result as above!

Because of the gaps in the data and the graph, we can't be completely sure we've found the best answer or that we've fully classified all points:

Depending on how expensive every foot of fencing is, we may choose to consider A as a function defined on the whole $[0, 50]$.

Let's review the tools at our disposal. Examining the graph reveals that the maximum value lies somewhere in the area where the graph is closest to the horizontal. In other words, this is where the slope of the *tangent line* is zero. But this number is the derivative of our function. From the *Power*

Formula, we have:

$$\frac{dA}{dW} = -2W + 50\,.$$

We set it equal to 0 and solve:

$$-2W + 50 = 0\,.$$

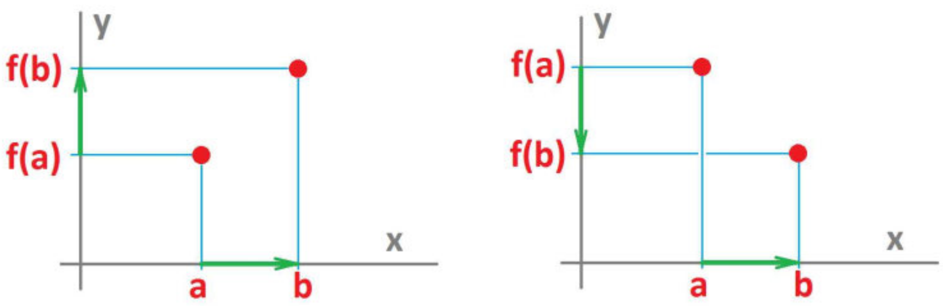
We conclude that at $W = 25$, the derivative is 0. And there are no other points like that.

Furthermore, examining the graph reveals that the function is increasing where the slopes of the tangent lines are positive, and decreasing where they are negative. In other words, the monotonicity is determined by the sign of the slope of the tangent line. We conclude that on the interval $(0, 25)$, the derivative is strictly positive and on $(25, 50)$ it is strictly negative. This analysis amounts to solving an inequality:

$$\frac{dA}{dW} > 0 \text{ when } W < 25 \quad\text{and}\quad \frac{dA}{dW} < 0 \text{ when } W > 25\,.$$

How can we find out about *any* given function whether and where it has monotonicity intervals and its max/min points? The answer we have discovered is *with the derivative* but we will reach this goal in several stages. First, some background.

We understand increasing functions as ones with graphs rising, and decreasing functions as ones with graphs falling. We also visualize monotonicity of functions in terms of parts of the graph above or below other parts. However, the precise definition must rely on comparing the values *two points at a time.*



Recall the definition (seen in Volume 1, Chapter 1PC-4):

Definition 5.1.2: increasing and decreasing functions

Suppose we have a function $y = f(x)$ and a subset I of its domain.

1. The function $y = f(x)$ is called *increasing on* I if, for all a, b in I, we have:

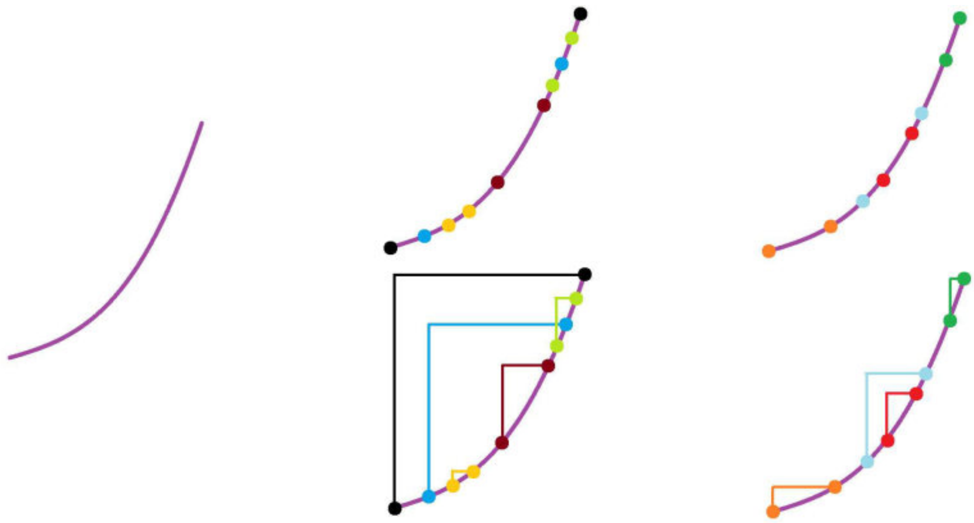
$$\text{if } a \le b\,,\ \text{then}\ \ f(a) \le f(b)\,.$$

2. The function is called *decreasing on* I if, for all a, b in I, we have:

$$\text{if } a \le b\,,\ \text{then}\ \ f(a) \ge f(b)\,.$$

The function is also called *strictly increasing* and *strictly decreasing* respectively if these pairs of values cannot be equal; i.e., we replace the non-strict inequality signs "$\le$" and "$\ge$" with strict "$<$" and "$>$". Collectively, these functions are called *monotone* and *strictly monotone.*

When the set I is an interval, the problem represents a significant challenge (as seen in Volume 1, Chapter 1PC-4) because drawing such conclusions means comparing *infinitely many* pairs of points:

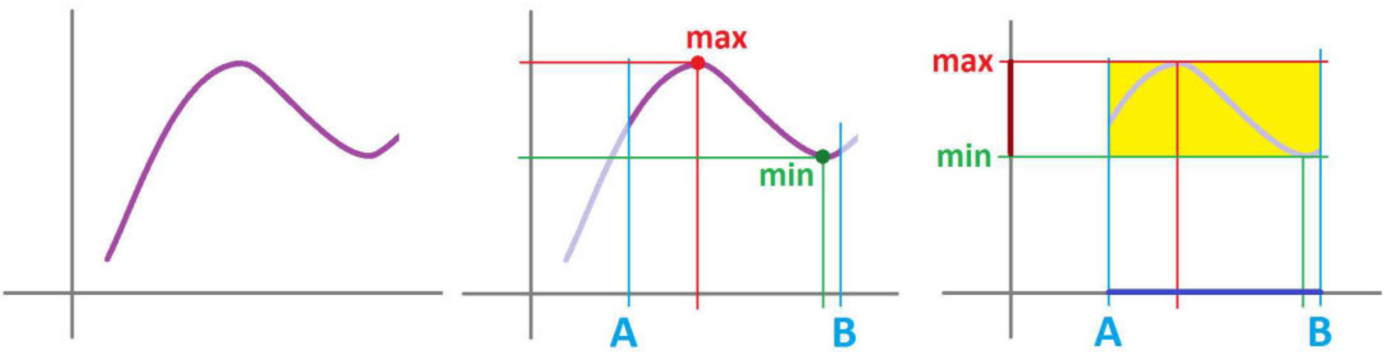


Example 5.1.3: transformations

What is the meaning of the definition if we look at functions as *transformations*? Let's place the x- and y-axis side by side:

An increasing function doesn't flip any parts of the interval and a decreasing flips all!

Next, maxima and minima:



Even though we understand the maxima and minima of functions as those locations on the graphs above or below all others, the precise definition must rely, once again, on comparing this location to *one point at a time*:

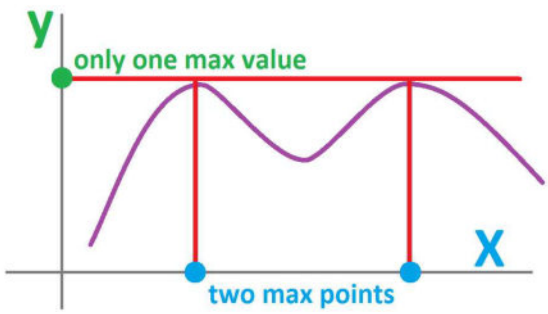
Definition 5.1.4: global maxima and minima

Suppose we have a function $y = f(x)$ and a subset I of its domain.

1. Number $x = c$ is called a *global maximum point* of f on I if $f(c)$ is the maximum value of the range of f on I, i.e.,
$$f(c) \geq f(x) \text{ for all } x \text{ in } I\,,$$
then $y = f(c)$ is called the *global maximum value* of f on I.
2. Number $x = c$ is called a *global minimum point* of f on I if $f(c)$ is the minimum value of the range of f on I, i.e.,
$$f(c) \leq f(x) \text{ for all } x \text{ in } I\,,$$

then $y = f(c)$ is called the *global minimum value* of f on I.
We call these *global extreme points and values*, or extrema.

As you can see, there can be many max *points* (those are x's) but only one max *value* (this is a y).

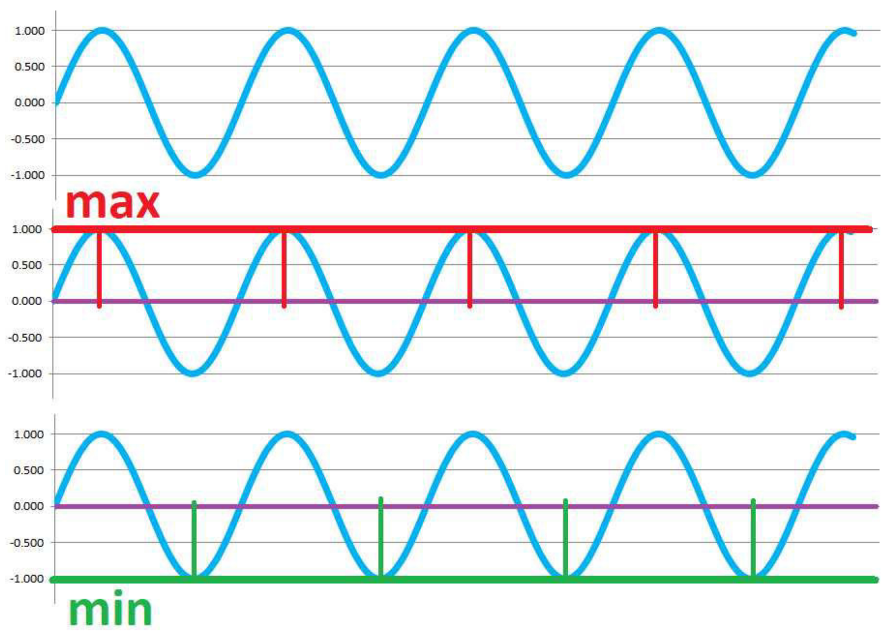


Indeed, there can be only one "highest elevation" in an area but there can be many "highest locations".

Example 5.1.5: many maximum points

Over the interval $I = (-\infty, +\infty)$, the function $f(x) = \sin x$ has only one global maximum *value*, $y = 1$, but many global maximum *points*,

$$x = \pi/2 + 2\pi k, \ k = 0, \pm 1, \pm 2, ...$$

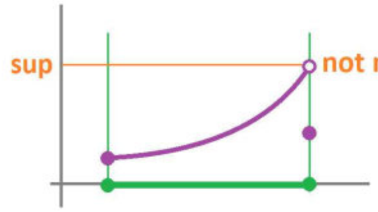


Similarly, the function has only one global minimum value, $y = -1$, but many global minimum points:

$$x = -\pi/2 + 2\pi k, \ k = 0, \pm 1, \pm 2, ...$$

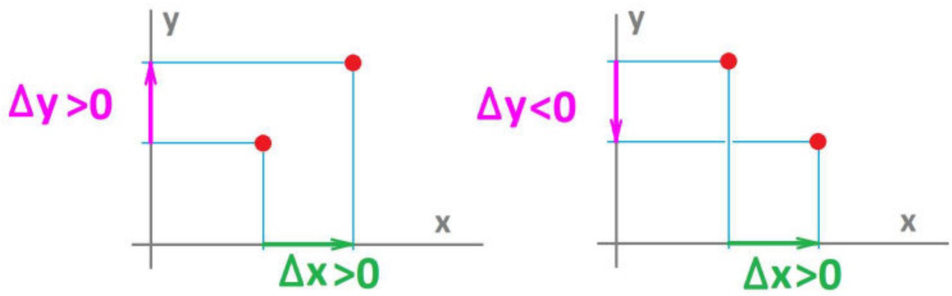
The function changes its monotonicity at its extreme points.

We will limit our attention to *continuous functions* in order to avoid the situation when these values – the supremum and the infimum – are never reached (Chapter 1):



The starting point of our analysis will be, as before, the differences Δf of f. Their *signs* determine whether

the function goes up or down from node to node of the partition:



The difference Δf and the difference quotient $\frac{\Delta f}{\Delta x}$ come hand in hand:

$$\boxed{f(x_k) \leq f(x_{k+1}) \iff \Delta f(c_k) \geq 0 \iff \frac{\Delta f}{\Delta x}(c_k) \geq 0}$$

The discrete case is solved! We write the summary below:

Theorem 5.1.6: Monotonicity from Sign of Difference

Suppose a function f is defined at the nodes of a partition of an interval. Then:

- *The function f is (strictly) increasing if and only if its difference Δf is non-negative (positive).*
- *The function f is (strictly) decreasing if and only if its difference Δf is non-positive (negative).*

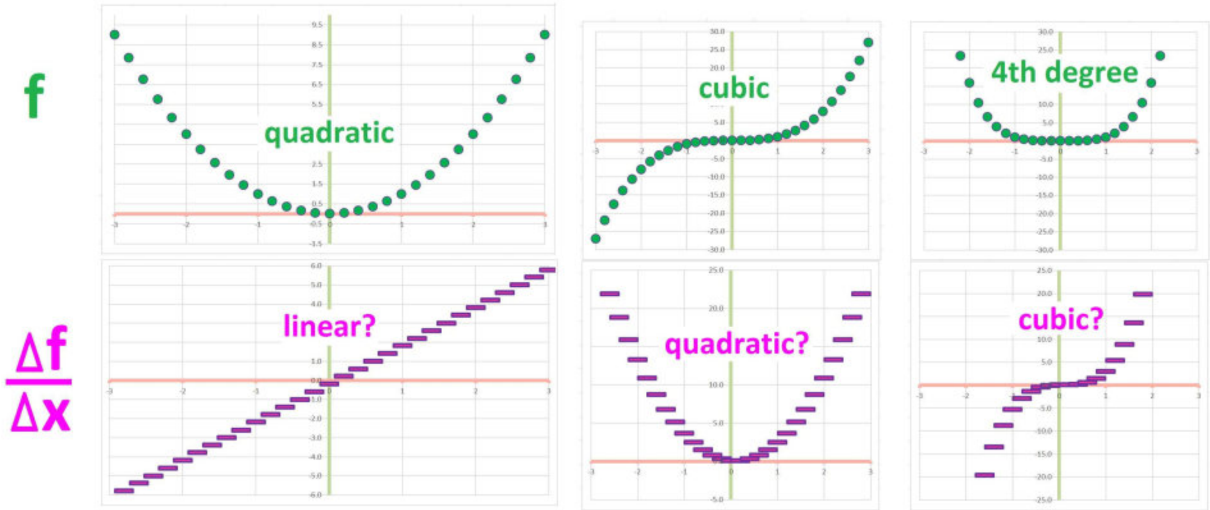
Now we just divide by $\Delta x > 0$:

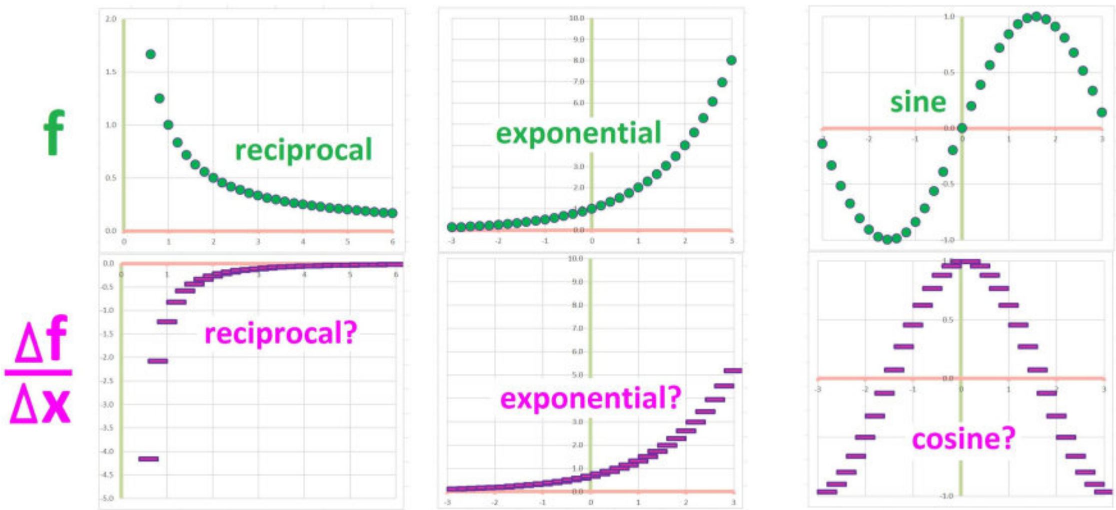
Theorem 5.1.7: Monotonicity from Sign of Difference Quotient

Suppose a function f is defined at the nodes of a partition of an interval. Then:

- *The function f is (strictly) increasing if and only if its difference quotient $\frac{\Delta f}{\Delta x}$ is non-negative (positive).*
- *The function f is (strictly) decreasing if and only if its difference quotient $\frac{\Delta f}{\Delta x}$ is non-positive (negative).*

The conclusions are confirmed by these examples:





How does this help with the study of monotonicity of functions defined on intervals, i.e., the "continuous" case? Taking the limit of the difference quotient – with a particular sign – will tell us about the *sign of the derivative.*

But first, let's review that we already know.

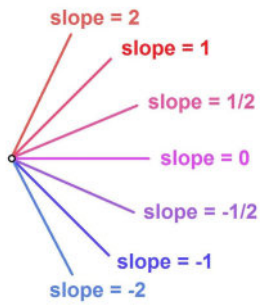
What do we know about a general linear function? Its derivative is its slope:

$$\begin{array}{ll} f(x) & = mx + b \\ f'(x) & = m \end{array}$$

So, we derive its monotonicity from the sign of its derivative by manipulating these inequalities (as seen in Volume 1, Chapter 1PC-5):

$$\begin{array}{llllll} m < 0 & \Longrightarrow & mu + b > mv + b & \text{if } u < v & \Longrightarrow & f \text{ is strictly decreasing} \\ m = 0 & \Longrightarrow & mu + b = mv + b & \text{if } u < v & \Longrightarrow & f \text{ is constant} \\ m > 0 & \Longrightarrow & mu + b < mv + b & \text{if } u < v & \Longrightarrow & f \text{ is strictly increasing} \end{array}$$

We have demonstrated that only the slopes matter:



This is the summary:

Theorem 5.1.8: Monotonicity of Linear Functions

A linear polynomial is:

1. *strictly increasing when its derivative is positive, and*
2. *strictly decreasing when its derivative is negative.*

Things become more complex if we need to analyze a quadratic polynomial:

$$f(x) = ax^2 + bx + c, \ a \neq 0 \,.$$

Recall that the graphs of all quadratic polynomials are *parabolas* (Volume 1, Chapter 1PC-4). In fact, they are all come from transformations of *the* parabola $y = x^2$. What matters especially, is the location of the vertex of the parabola:

$$v = -\frac{b}{2a} \,.$$

The reason is that the line $x = v$ is the axis of symmetry of this parabola.

Then, we conclude:

- If $a > 0$, then f is strictly decreasing on $(-\infty, v)$ and strictly increasing on $(v, +\infty)$.
- If $a < 0$, then f is strictly increasing on $(-\infty, v)$ and strictly decreasing on $(v, +\infty)$.

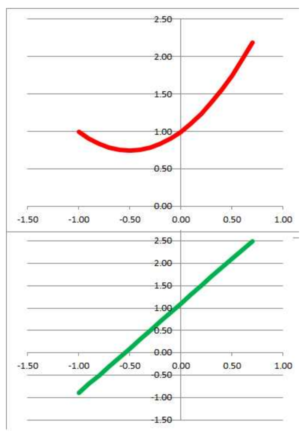
But, we will also discover, the vertex is where the derivative *changes its sign*! Indeed, its derivative,

$$f'(x) = 2ax + b\,,$$

is a linear function, and we easily derive the following:

- If $a > 0$, then $f' < 0$ on $(-\infty, v)$ and $f' > 0$ on $(v, +\infty)$.
- If $a < 0$, then $f' > 0$ on $(-\infty, v)$ and $f' < 0$ on $(v, +\infty)$.

This is the link:



Below is the conclusion:

Theorem 5.1.9: Monotonicity of Quadratic Functions

A quadratic function is:

1. *strictly increasing when its derivative is positive, and*
2. *strictly decreasing when its derivative is negative.*

The two theorems match!

Example 5.1.10: rectangular enclosure

For the first example of this section, the vertex of the parabola is at

$$v = \frac{0+50}{2} = 25\,.$$

We then derive the *dynamics* of this situation:

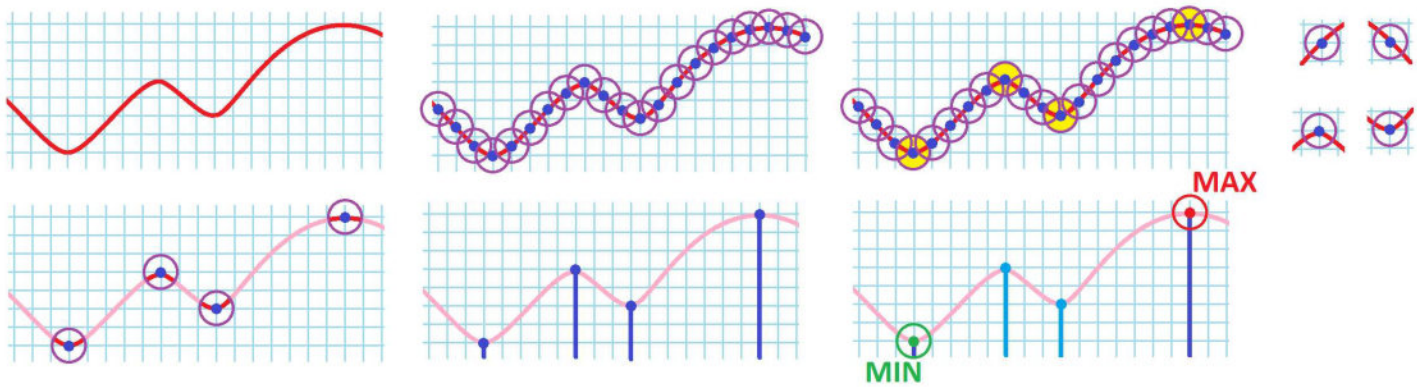
- As we increase the width from 0, the area also increases.
- As the width reaches 25, the area reaches its maximum value of 625.
- As we increase the width past 25, the area starts to decrease.

With more and more complex functions, the analysis becomes more and more challenging.

We will take an indirect approach: *from local to global.* Even though our interest is the monotonicity of functions on whole intervals and the global extrema, we will first solve the local problems:

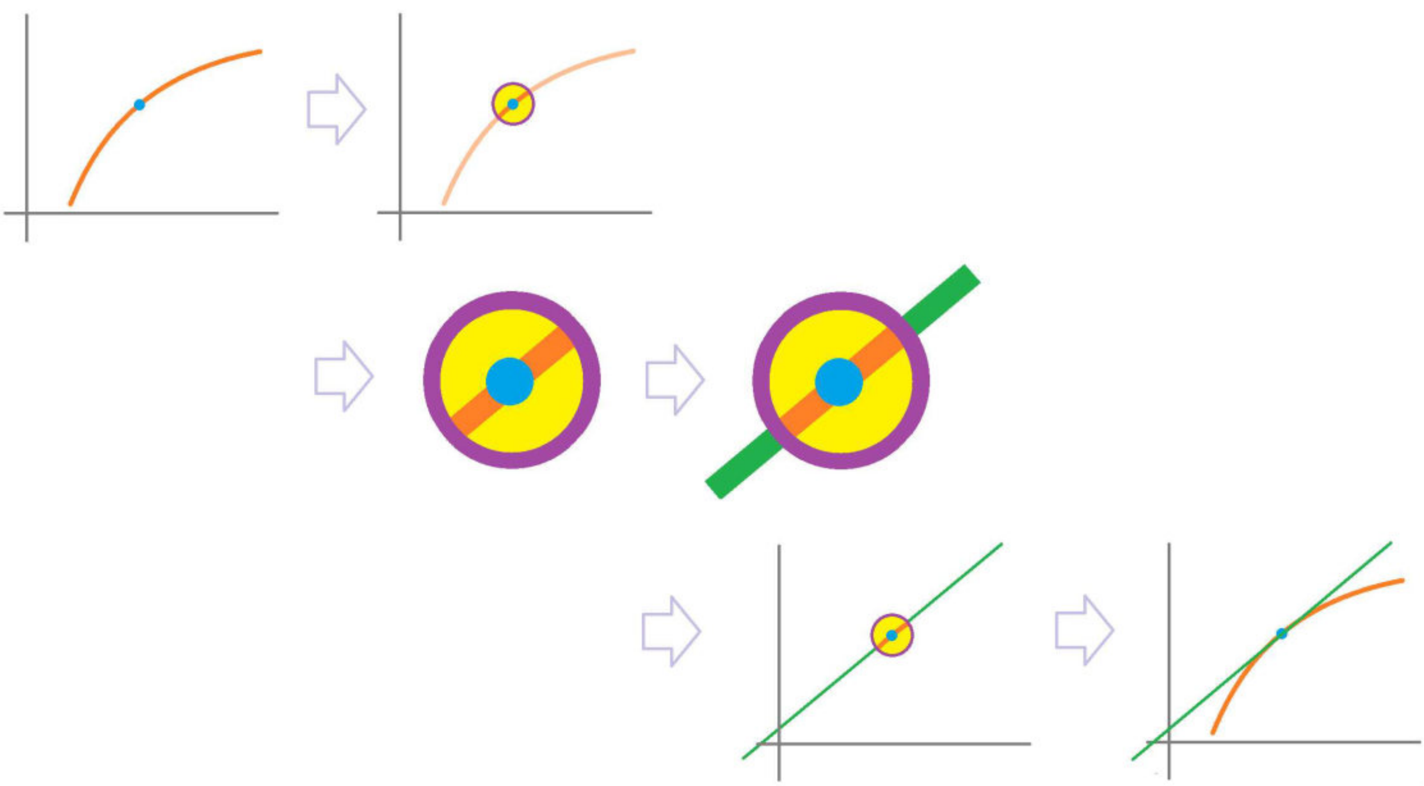
1. Is the function strictly monotone in the *vicinity* of a point?
2. Is the value of the function at a particular point the largest or the smallest in the *vicinity* of the point?

In our search for the extrema, the former are the losers and the latter are the potential winners of this contest:



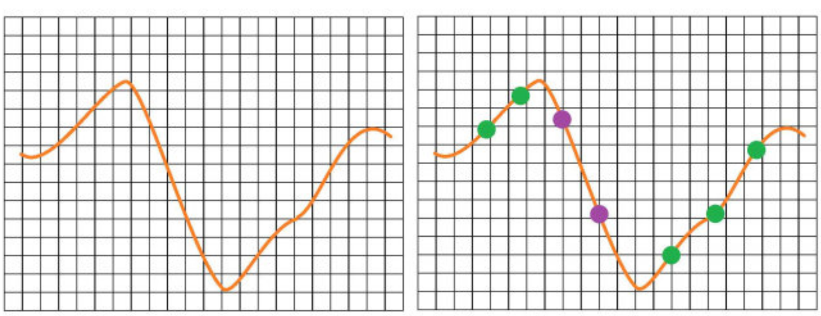
We will now learn how to use the *derivative* to find the intervals of monotonicity and the extreme points.

The reason for our approach is that the information about the function's behavior that the derivative encodes is *local*. Indeed, no matter how small a piece of the graph around the point $(a, f(a))$ you keep, the derivative $f'(a)$ at that point will remain the same.

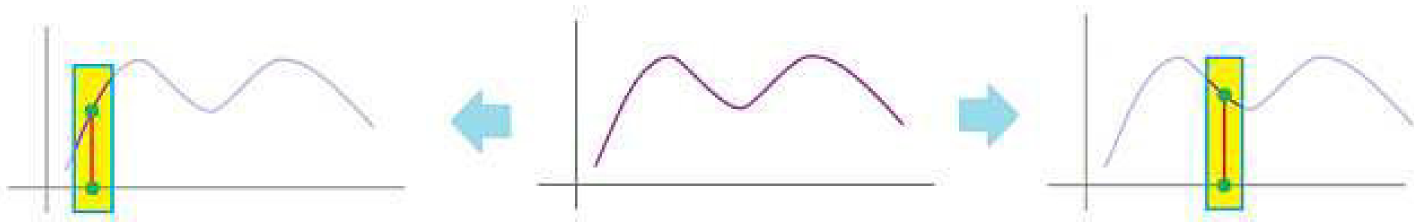


The two definitions below are the stepping stones toward a realistic method for finding what is defined above.

The global analysis of monotonicity is supposed to supply us with the intervals where the function is monotone:



With the local approach, in contrast, only the behavior in the vicinity of the point matters:



Let's make it precise:

Definition 5.1.11: locally increasing and decreasing

1. A function f is *locally increasing* at $x = c$ if for all x in some open interval I within the domain of f containing c, we have:
   - $f(x) \leq f(c)$ for all $x < c$, and
   - $f(x) \geq f(c)$ for all $x > c$.
2. A function f is *locally decreasing* at $x = c$ if for all x in some open interval I within the domain of f containing c, we have:
   - $f(x) \geq f(c)$ for all $x < c$, and
   - $f(x) \leq f(c)$ for all $x > c$.

We also say that f is *locally monotone* at these points.

In other words, we have:

	$x <$	c	$< x$
$f \nearrow$	$f(x) \leq$	$f(c)$	$\leq f(x)$
$f \searrow$	$f(x) \geq$	$f(c)$	$\geq f(x)$

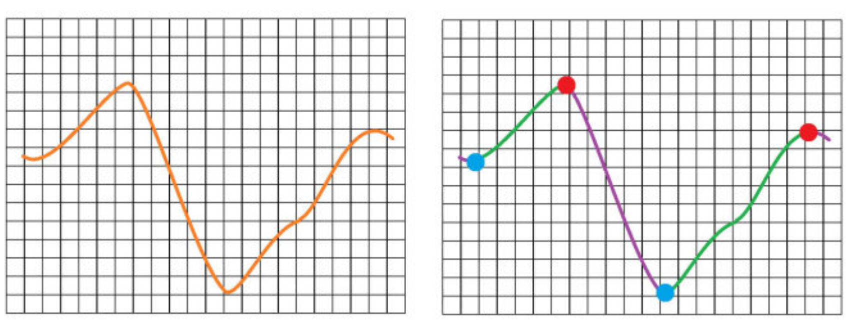
Exercise 5.1.12

Does the definition mean that there is an open interval I around c such that f restricted to I is increasing or decreasing, respectively?

Next, the *extreme points.*

First, monotone points are *not* among the candidates!

The global analysis of monotonicity is supposed to supply us with the highest point over whole intervals:



With the local approach, in contrast, only the behavior in the vicinity of a point matters:

Let's make it precise:

Definition 5.1.13: local minimum and maximum points

1. A function f has a *local minimum point* at $x = c$ if for all x in some open interval I that contains c, we have:
$$f(c) \le f(x)\,.$$
2. A function f has a *local maximum point* at $x = c$ if for all x in some interval I that contains c, we have:
$$f(c) \ge f(x)\,.$$

We call these *local extreme points*, or local extrema.

In other words, there is an open interval I around c such that c is the global maximum (or minimum) point when f is restricted to I.

Warning!

The inequalities in both definitions are non-strict.

These *are* our candidates for the global extrema!

A summary of the two concepts defined above is illustrated below:

The point of interest on the graph is red, while the rest of the graph is to be located somewhere in the pink area.

We have the following classification of points:

Theorem 5.1.14: Local Monotonicity and Extrema

At every point of its domain, a function cannot both be locally monotone and have a local extremum.

Exercise 5.1.15

Give an example of a function with points where it neither is locally monotone nor has a local extremum.

The connection of these two – local – concepts to the previous – global – concepts is as follows:

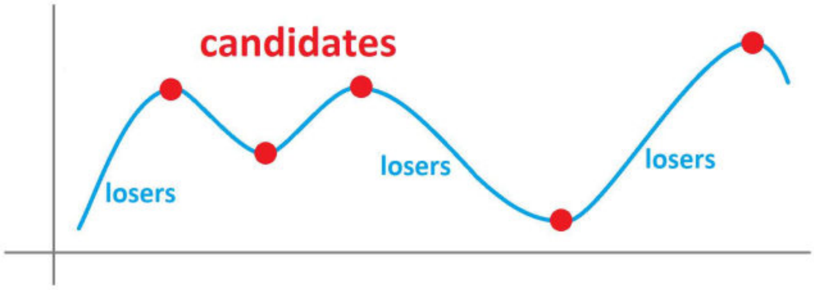
Theorem 5.1.16: From Local to Global Monotonicity

If a function is increasing (or decreasing) on an open interval I, then it is locally increasing (or decreasing) at all points in I.

Theorem 5.1.17: From Local to Global Extrema

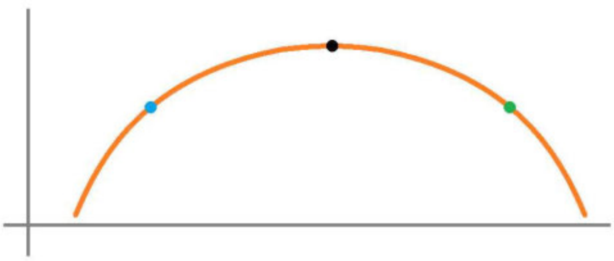
If a function has a global maximum (or minimum) point at $x = c$, on an open interval I, then $x = c$ is a local maximum (or minimum).

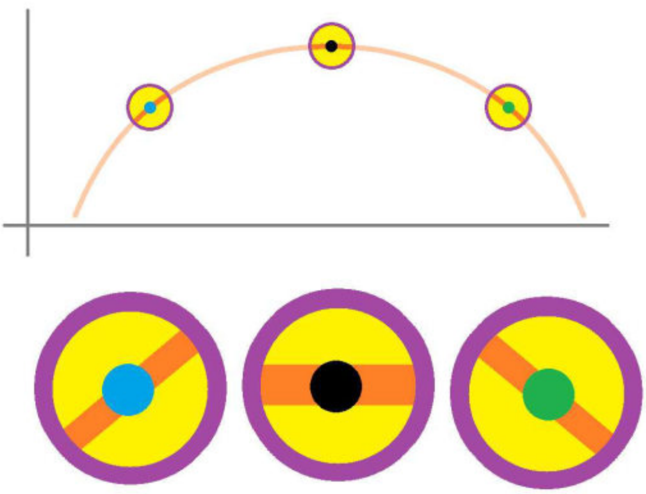
In the first stage of the contest for the global extrema, we eliminate infinitely many participants:



We only keep the local extrema. Typically, we are left with just a few.

Now, how do we find all these points? We look at the slopes:





The three parts of the last picture suggest the following respectively:

1. If the function is locally increasing at a point c, its graph would have an increasing tangent line, i.e., $f'(c) \geq 0$.
2. If the function has a local extreme point at c, its graph would have a horizontal tangent line, i.e., $f'(c) = 0$.
3. If the function is locally decreasing at a point c, its graph would have a decreasing tangent line, i.e., $f'(c) \leq 0$.

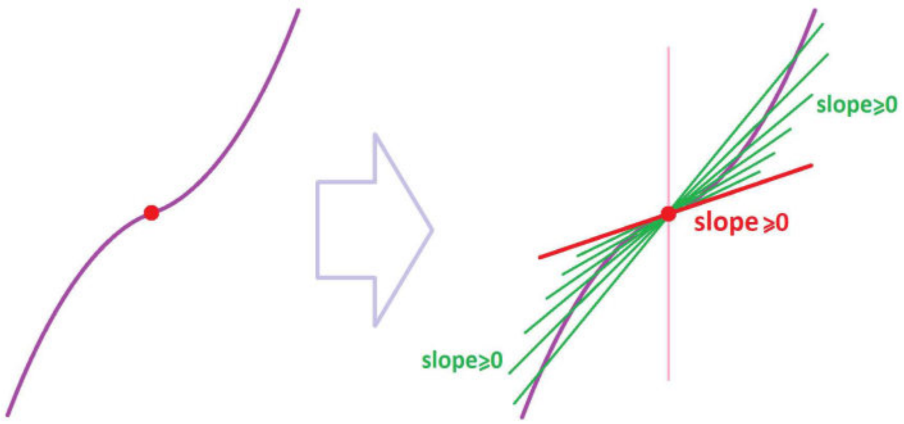
The reasoning for the second is that if $f'(c) \neq 0$, we have either $f'(c) > 0$ or $f'(c) < 0$. If we zoom in, the graph merges into the tangent line and we realize there can be no max or min!

Theorem 5.1.18: Local Monotonicity

Suppose a function $y = f(x)$ is differentiable at $x = c$. Then if f is locally increasing or decreasing at point $x = c$, then, respectively,

$$f'(c) \geq 0 \quad \texttt{OR} \quad f'(c) \leq 0\,.$$

The idea of the proof is to watch for the convergence of the secant lines to the tangent line:



Proof.

Let's suppose that f is locally increasing at c. Now, $f'(c)$ is the limit of the slopes of the secant lines through the point $(c, f(c))$ as shown above.

Consider secant lines within the interval. Then the secant lines *both* to the left and to the right of c have non-negative slopes:

$$\text{slope } = \frac{\overbrace{f(c) - f(x)}^{\geq 0}}{\underbrace{c - x}_{\geq 0}} \geq 0\,.$$

Apply now the *Comparison Theorem* for limits (Chapter 1):

$$g(x) \geq 0 \implies \lim_{x \to c} g(x) \geq 0\,.$$

It follows that $f'(c) \geq 0$.

Exercise 5.1.19

Based on the theorem, finish the sentence "If $f'(c) \geq 0$, then ...."

Exercise 5.1.20

Based on the theorem, finish the sentence "If $f'(c) < 0$, then ...."

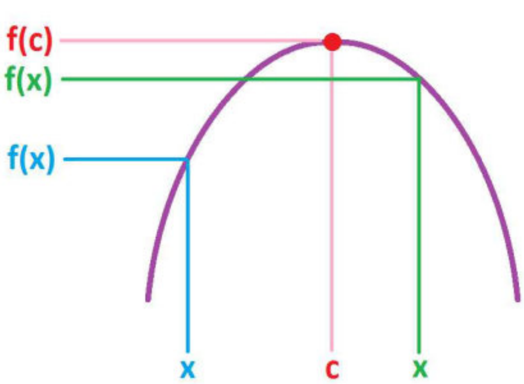
The following is crucial for optimization problems:

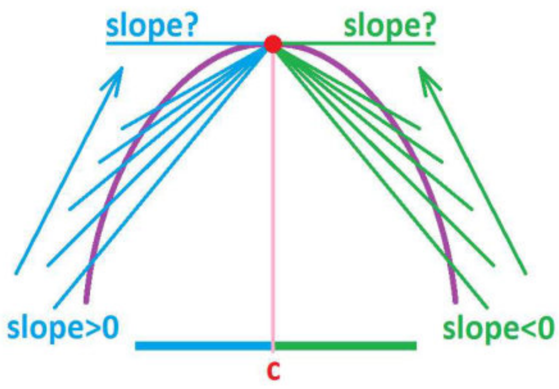
Theorem 5.1.21: Local Extrema (Fermat's Theorem)

Suppose a function $y = f(x)$ is differentiable at $x = c$. Then if $x = c$ is a local extreme point of $y = f(x)$, then

$$f'(c) = 0\,.$$

The idea of the proof is, again, to watch for the convergence of the secant lines to the tangent line:





Proof.

Let's suppose c is a local *maximum*: $f(c) \geq f(x)$ within some open interval that contains c. Again, $f'(c)$ is the limit of the slopes of the secant lines through the point $(c, f(c))$. The idea of the proof above is applied but separately for the points to the left and to the right of c.

Consider secant lines within the interval.

$x <$	c	$> x$
slopes ≥ 0		slopes ≤ 0
therefore		therefore
$f'(c) \geq 0$	`AND`	$f'(c) \leq 0$
$0 \leq$	$f'(c)$	≤ 0
therefore		therefore
$0 =$	$f'(c)$	$= 0$

For completeness, we demonstrate algebraically that the slopes of these secants have these signs. First, take any (secant) line through $(x, f(x))$ and $(c, f(c))$ with $x < c$. Then $f(x) \leq f(c)$ when x is close enough to c. Why? Because c is a local max (review the definition). Then we have:

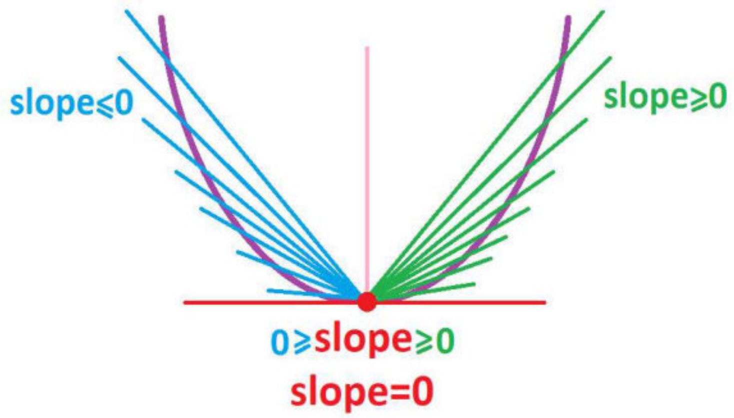
$$\text{slope} \ = \frac{\overbrace{f(c) - f(x)}^{\geq 0}}{\underbrace{c - x}_{>0}} \geq 0\,.$$

Second, take any (secant) line through $(x, f(x))$ and $(c, f(c))$ with $x > c$. Then $f(x) \leq f(c)$ when x is close enough to c, because c is a local max. Then we have:

$$\text{slope} \ = \frac{\overbrace{f(c) - f(x)}^{\geq 0}}{\underbrace{c - x}_{<0}} \leq 0\,.$$

The *Comparison Theorem* for limits is applied, just as in the last proof.

The proof for a *minimum* is similar. Here is a summary:



Exercise 5.1.22

Based on the theorem, finish the sentence "If $f'(c) \neq 0$, then ...."

Exercise 5.1.23

Based on the theorem, finish the sentence "If $f'(c) = 0$, then ...."

Warning!

In this context, the derivative is just a means to an end.

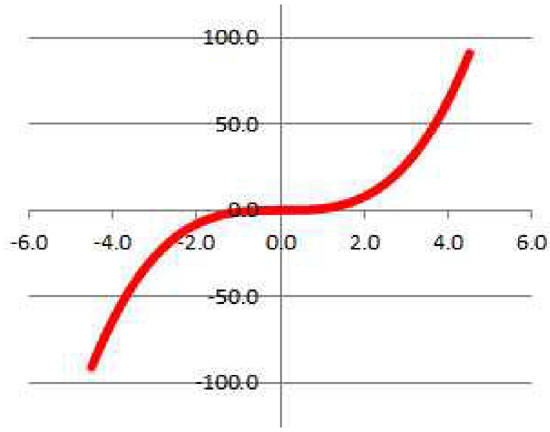
Now what about vice versa? Is the *converse* of Fermat's Theorem true, that:

- $f'(c) = 0$, then c is a local extreme point of f?

It's *not* true, as the example below shows.

Example 5.1.24: zero derivative of x^3

Consider $f(x) = x^3$ at $x = 0$:



We have:

$$f'(x) = 3x^2 \implies f'(0) = 0\,,$$

but this is not an extreme point. The function is increasing everywhere!

Exercise 5.1.25

How about a "strict" Local Monotonicity Theorem: If f is locally strictly increasing or strictly decreasing at $x = c$, then, respectively,

$$f'(c) > 0 \;\text{ or }\; f'(c) < 0\,.$$

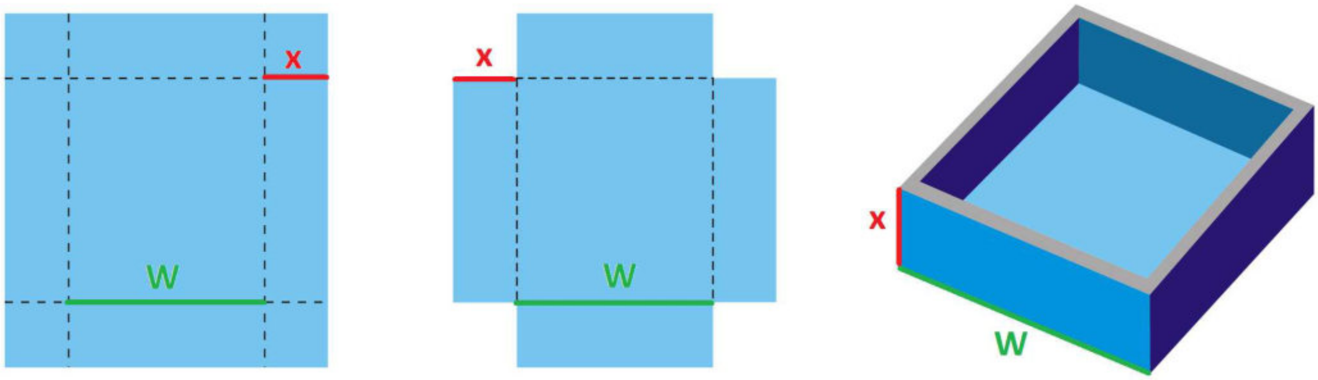
Warning!

An alternative terminology to use is:

- "absolute" extreme points instead of "global", and
- "relative" extreme points instead of "local".

Example 5.1.26: maximize open box

Let's consider a different optimization problem: The corners are to be cut from a 10×10 piece of cardboard to create an *open* box of the largest possible volume.

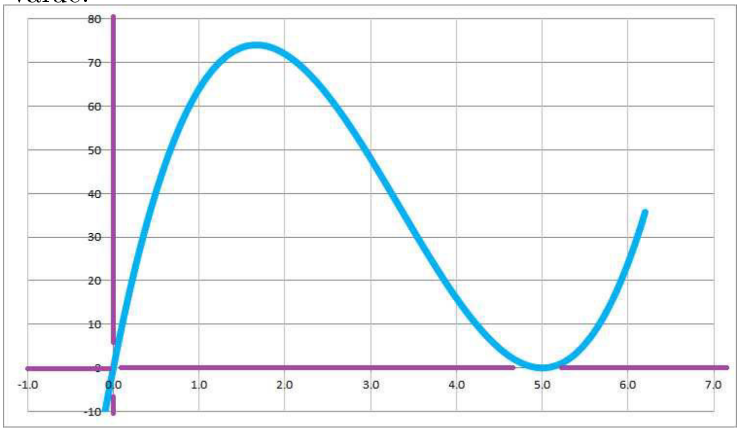


Let's denote the side of the little square by x. Then x becomes the height of the box with the width equal to $10 - 2x$. Then the volume of the box is

$$V = x(10-2x)^2 = 4x^3 - 40x^2 + 100x\,.$$

The function is cubic! We need to find the largest possible value of this function (for $0 \le x \le 5$), but, unfortunately, the completeness of information about the quadratic functions isn't matched by what we know about these.

If we plot its graph, we see the highest point within this interval, but, without symmetry to rely on, we can't know its *exact* value:



We know, however, from *Fermat's Theorem* that we can find this point as one with a zero derivative; i.e., this x satisfies the equation:

$$V' = (4x^3 - 40x^2 + 100x)' = 12x^2 - 80x + 100 = 4(3x^2 - 20x + 25) = 0\,.$$

Then, from the *Quadratic Formula*, we have:

$$x = \frac{20 \pm \sqrt{20^2 - 4\cdot 3\cdot 25}}{2\cdot 3} = \frac{20 \pm 10}{6} = 5,\ \frac{5}{3}\,.$$

There are no other candidates for this max point! Therefore, the latter one is the answer. We can also confirm that all points between the two are increasing according to the *Local Monotonicity Theorem*

by solving this inequality:

$$V' = 4(3x^2 - 20x + 25) > 0\,.$$

Indeed, we have the two roots, 5 and $\frac{5}{3}$, and between them V' will remain negative.

This is how the monotonicity problem is solved so far:

$$\boxed{f(x_k) \le f(x_{k+1}) \iff \Delta f(c_k) \ge 0 \iff \frac{\Delta f}{\Delta x}(c_k) \ge 0 \implies \frac{df}{dx} \ge 0}$$

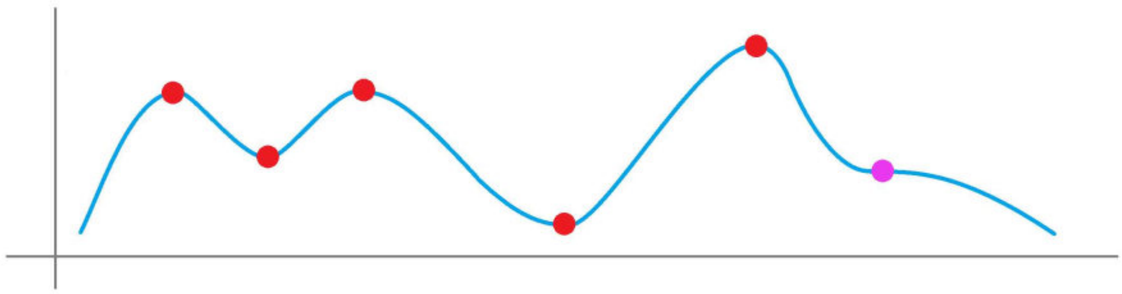
The last arrow goes in one direction!

5.2. Optimization of functions

According to *Fermat's Theorem*, the points with a zero derivative include all local extrema as well as some other points. We add all of those to our list:

- ▶ The points with a zero derivative are *candidates* for global extreme points.

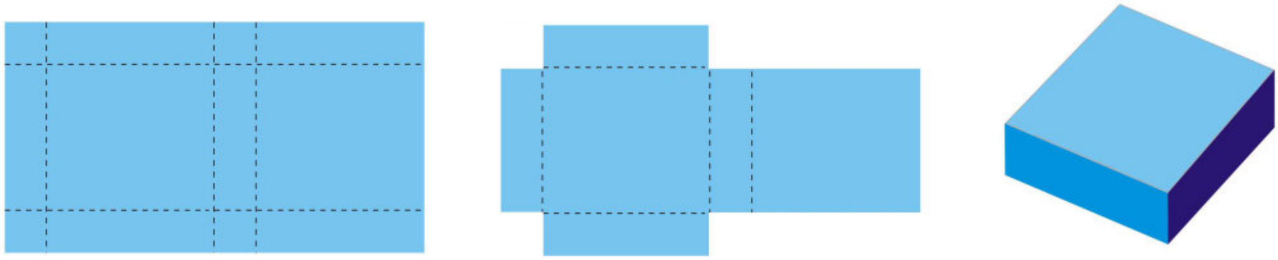
For example, below we have six candidates:



The blue points have been already eliminated from the contest. The six corresponding values of the function are still to be compared to find the final winner or winners.

Example 5.2.1: maximize closed box

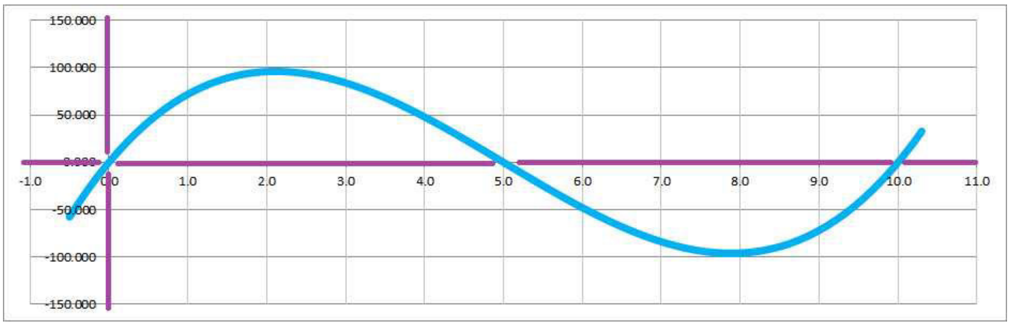
Let's modify our optimization problem: The corners are to be cut from a 10×20 piece of cardboard to create a *closed* box of the largest possible volume.



Let y be the width of the box; then $2x + 2y = 20$, or $y = 10 - x$. The volume of the box is then:

$$V = x(10 - 2x)(10 - x) = 2x^3 - 40x^2 + 100x\,,$$

where x is the side of the little square under the restriction $0 \le x \le 5$. Plotting the graph (with a spreadsheet) suggests that there is indeed a local maximum:



What is its exact value? According to *Fermat's Theorem*, since the function is differentiable, the point, c, has to satisfy $f'(c) = 0$. Find the derivative:

$$V'(x) = (2x^3 - 30x^2 + 100x)' = 6x^2 - 60x + 100\,.$$

Then solve this equation:

$$V'(x) = 6x^2 - 60x + 100 = 0\,,$$

or

$$3x^2 - 30x + 50 = 0\,.$$

By the *Quadratic Formula*, we have:

$$c = \frac{30 \pm \sqrt{30^2 - 4 \cdot 3 \cdot 50}}{2 \cdot 3} = \frac{30 \pm 10\sqrt{3}}{6}\,.$$

The smaller answer, $c \approx 2.1$, is the maximum.

Example 5.2.2: extrema of $\sin$

Consider $f(x) = \sin x$. We follow the same plan. First, the derivative is found and set equal to 0:

$$\frac{d}{dx}(\sin x) = \cos x = 0\,.$$

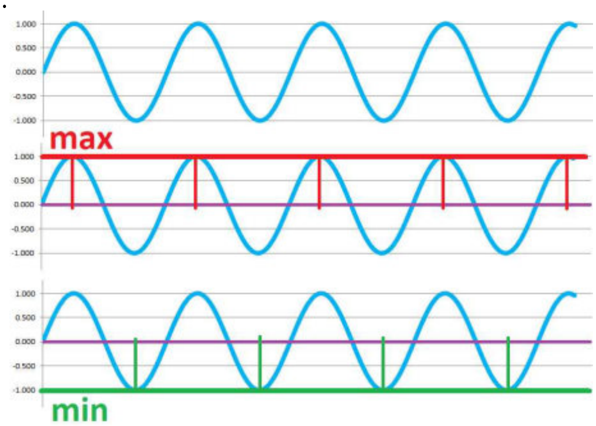
The equation produces the same list of candidates:

$$x = k\pi + \pi/2,\ \ k = 0, \pm 1, \pm 2, \ldots$$

Just from this fact alone, we can't tell which ones are maxima and which ones are minima. However, we know the following:

$$\sin\left(k\pi + \pi/2\right) = \begin{cases} -1 & \text{if } k \text{ is odd,} \\ 1 & \text{if } k \text{ is even.} \end{cases}$$

Therefore, the former are the minima and the latter are the maxima. This conclusion confirms what we know about this function:



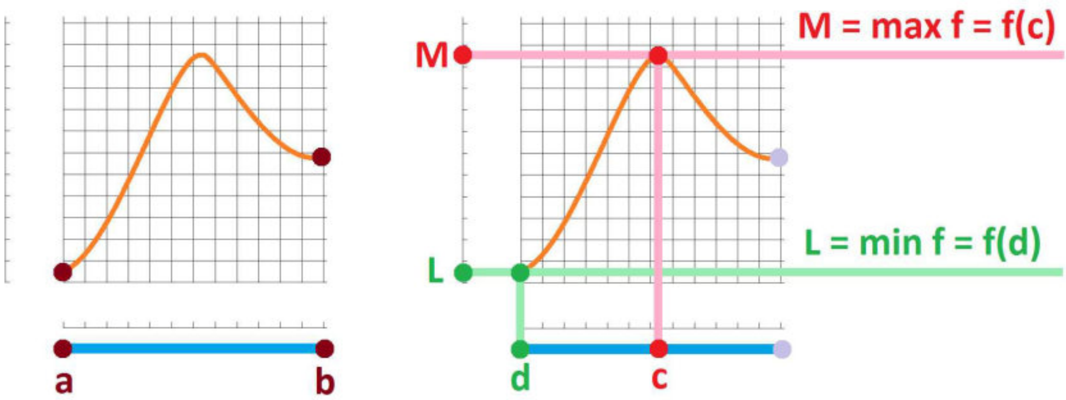
Can there be other candidates for global extrema?

When the problem calls for limiting our attention to a *closed interval*, its end-points cannot be local extrema because the function is only defined on one side of such a point. We simply add the two to our list:

▶ The end-points of the interval are *candidates* for global extreme points.

For example, below if we ignored the right-most point, we'd give up the global minimum:

According to the *Extreme Value Theorem*, a continuous function on a bounded closed interval has a global maximum and a global minimum:



We conclude:

▶ The list of all points of zero derivative with the end-points added will always contain all extreme points.

This is the justification of our strategy.

Stripped of all the incidental details of a word problem, this is what a solution of an optimization problem will look like.

Example 5.2.3: optimize on closed interval

Find global extreme points of

$$f(x) = x^3 - 3x \quad \text{on } [-2, 3]\,.$$

Step 1: Find the points with zero derivative. Compute:

$$f'(x) = 3x^2 - 3\,.$$

Set it equal to 0 and solve for x.

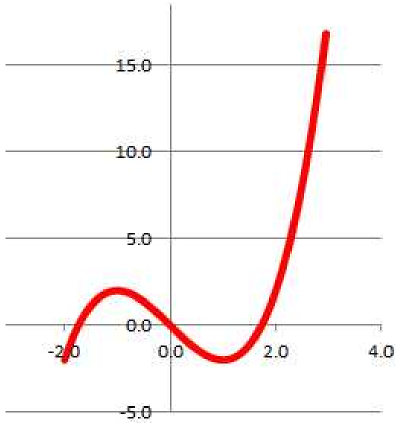
$$\begin{aligned} 3x^2 - 3 &= 0 \quad \Longrightarrow \\ x^2 &= 1 \quad \Longrightarrow \\ x &= \pm 1\,. \end{aligned}$$

Step 2: Compare the values of f at these points and at the end-points of the interval; find the smallest

and the largest:

Candidates	x	$f(x)$	$= x^3 - 3x = x(x^2-3)$		values	classification
$f'(x) = 0$	1	$f(1)$	$= 1(1^2 - 3)$	$=$	-2	global min
$f'(x) = 0$	-1	$f(1)$	$= -1((-1)^2 - 3)$	$=$	2	possibly local but not global max
a	-2	$f(-2)$	$= -2((-2)^2 - 3)$	$=$	-2	global min
b	3	$f(3)$	$= 3(3^2 - 3)$	$=$	18	global max

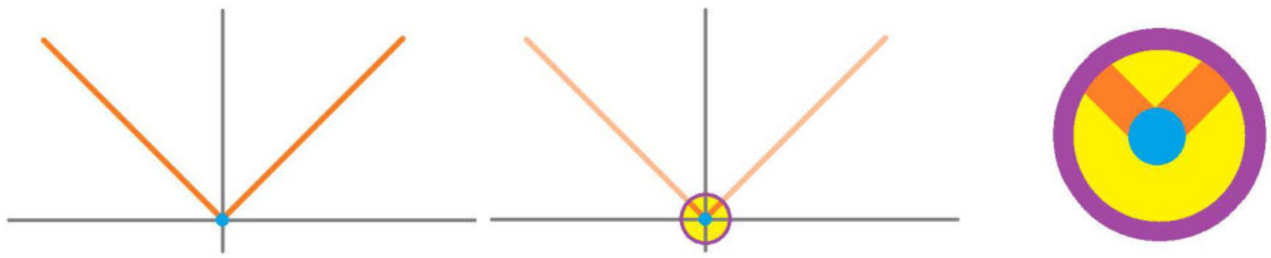
These points and how they are classified are visible on the graph:



According to Fermat's Theorem, there are no other candidates. The theorem, however, leaves an option of a *non-differentiable* function.

Example 5.2.4: non-differentiable

The absolute value function $f(x) = |x|$ has its global minimum at $x = 0$; however, it's not differentiable:



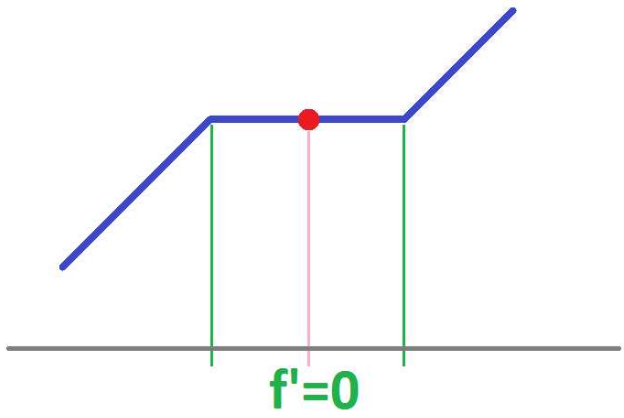
In fact, we have:

$$f'(x) = \begin{cases} -1 & \text{if } x < 0, \\ \text{undefined} & \text{if } x = 0, \\ 1 & \text{if } x > 0. \end{cases}$$

Therefore, the global minimum on any interval $[a, b]$ with $a < 0 < b$ is at $x = 0$, and the global maximum at a or b.

Exercise 5.2.5

Consider the derivative and the extreme points of this function:



We may have to add those to our list too:

▶ The points with undefined derivative are *candidates* for global extreme points.

The points with either undefined or zero derivative are often called "critical points".

Thus, our list of values of x to be checked is comprised of just two types: the critical points and the end-points. Then, from this list, we find:

- The x's with the largest value of y are the global maxima.
- The x's with the smallest value of y are the global minima.

Example 5.2.6: cubic polynomial

Analyze the function
$$f(x) = 2x^3 - 3x^2 - 12x + 1$$
on $[-2, 3]$.

Compute the derivative:
$$f'(x) = 6x^2 - 6x - 12\,.$$
Set it equal to zero and solve the equation:
$$\begin{array}{lll} 6x^2 - 6x - 12 & = 0 & \Longrightarrow \\ x^2 - x - 2 & = 0 & \Longrightarrow \\ \text{QF:} \qquad x & = \dfrac{1 \pm \sqrt{1-(-2)4}}{2} & \\ & = \dfrac{1 \pm 3}{2} = 2,\ 1\,. & \end{array}$$

Compare the values:

Candidates	x	$f(x) = 6x^2 - 6x - 12$	y
$f'(x) = 0$	2	$2 \cdot 2^3 - 3 \cdot 2^2 - 12 \cdot 2 + 1 = 16 - 12 - 24 + 1 =$	-19
$f'(x) = 0$	-1	$2 \cdot (-1)^3 - 3 \cdot (-1)^2 - 12 \cdot (-1) + 1 = -2 - 3 + 12 + 1 =$	8
a	-2	$2 \cdot (-2)^3 - 3 \cdot (-2)^2 - 12 \cdot (-2) + 1 = -16 - 12 + 24 + 1 =$	-3
b	3	$2 \cdot 3^3 - 3 \cdot 3^2 - 12 \cdot 3 + 1 = 54 - 27 - 36 + 1 =$	-8

Answer:
- The global max *value* is $y = 8$ attained at $x = -1$, a global max *point*.
- The global min *value* is $y = -19$ attained at $x = 2$, a global min *point*.

Warning!

Substitute the values into the original function, not the derivative.

This is a summary of our method:

Theorem 5.2.7: Global Extrema via Derivative

Suppose f is continuous on $[a,b]$ and c in $[a,b]$ is a global extreme point. Then, one of the following must be satisfied:

1. $f'(c)=0$, *or*
2. $f'(c)$ *is undefined, or*
3. $c=a$ *or* $c=b$.

Exercise 5.2.8

Redo the two examples about maximizing the volume of the box by following this method.

5.3. *What the derivative says about the difference quotient: The Mean Value Theorem*

The *Fermat's Theorem* is an example of a theorem the converse of which isn't true. This is what it does and does not say about a function differentiable at $x=c$:

$$\begin{array}{c|l|c}
x=c \ \text{ is} & \begin{array}{l}\text{a local}\\ \text{extreme point}\end{array} & \begin{array}{c}\Longrightarrow\\ \not\Longleftarrow\end{array} \quad \boxed{f'(c)=0}
\end{array}$$

In contrast, the converse of the *Local Monotonicity Theorem* is true but hasn't been proven yet:

$$\begin{array}{c|l|c}
x=c \ \text{ is} & \begin{array}{l}\text{a local}\\ \text{monotone point}\end{array} & \begin{array}{c}\Longrightarrow\\ \Longleftarrow\end{array} \quad \boxed{f'(c)\geq 0}
\end{array}$$

Recall from the last chapter, that there are other statements in the converses of which we are interested:

1. The derivative of a constant function is zero, but are the constants the only functions with this property?
2. The derivative of a linear polynomial is constant, but are the linear polynomials the only functions with this property?
3. The derivative of a quadratic polynomial is linear, but are the quadratic polynomials the only functions with this property?

Exercise 5.3.1

The first question asks if we can build a curve – other than a horizontal line – with all tangent lines horizontal. Try it!

Example 5.3.2: free fall

Answering Yes to the second and third questions solves the problem of *free fall* as presented in Chapter

4:

constant force $\Longrightarrow$ constant acceleration $\Longrightarrow$ linear velocity $\Longrightarrow$ quadratic position function

We have justified, therefore, our formula:

$$y(t) = y_0 + v_0 t - \frac{1}{2} g t^2 \,.$$

Furthermore, the trajectory of a ball thrown forward is a parabola!

The main theorem of this section will help with these and other questions.

Example 5.3.3: driving

Let's interpret the conditions of Fermat's Theorem in terms of motion:
- x is time.
- $f(x)$ is the location at time x.
- $f'(x)$ is the velocity at time x.

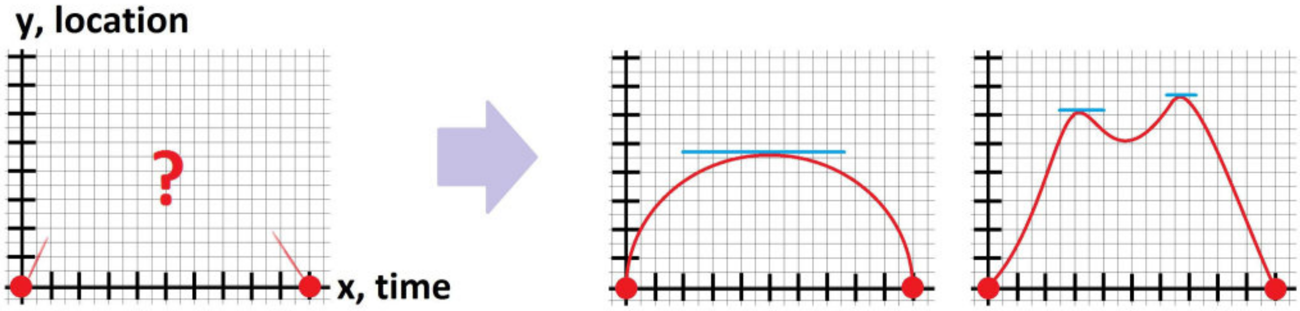
So, we assume that the velocity always makes sense; i.e., there are no sudden changes of direction, bumps, or crushes (and no teleportation!).

Now we interpret the conclusion of Fermat's Theorem in terms of motion:
- ▶ ($\Rightarrow$) Whenever we are the *farthest* from home or any location, we stop, at least for an instant.

But not conversely:
- ▶ ($\not\Leftarrow$) Even if we stop, we can resume driving in the same direction.

Imagine that you:
- left home at 1 pm,
- did some driving, and then
- came home at 2 pm.

Question: What can someone who stayed home infer about your *speed* during this time? For simplicity, we assume that you drove on a straight road. Then this is what that person knows (left):



The possibilities are numerous (right). You may have driven slowly, then fast, but one thing is certain: you came back home. And to come back, you had to turn around. And to turn around, you had to stop. So, *speed was zero* at least once! There is no knowledge *when* this happened though.

Let's make the observation made in the last example purely mathematical and turn it into a theorem:

Theorem 5.3.4: Rolle's Theorem

Suppose we have a function:
1. *f is continuous on $[a, b]$.*
2. *f is differentiable on (a, b).*

Then, if

$$f(a) = f(b) \,,$$

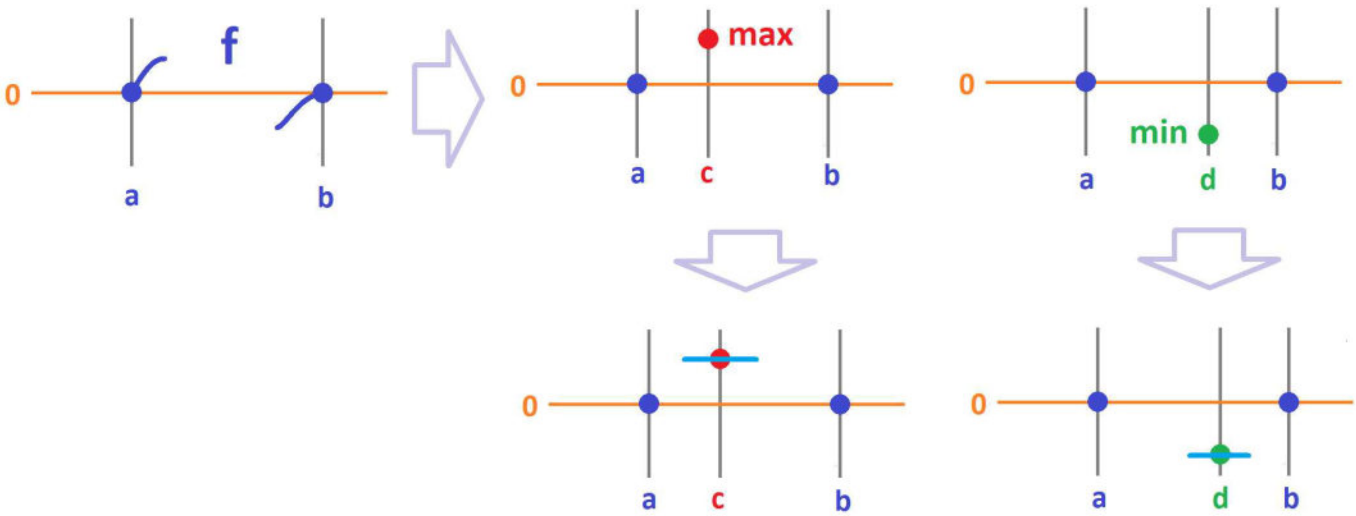
we have:

$$f'(c) = 0$$

`FOR SOME` c in (a, b) .

We again assume that x is time, limited to interval $[a, b]$. In the theorem, #1 means that you don't leap and #2 means that you don't crash. We also assume that you've come back. We already know that this special point in time was when you were farthest away from home (in either direction); you weren't moving then.

The idea of the proof is below:



Proof.

Suppose f has on $[a, b]$:

- a global maximum at $x = c$ and
- a global minimum at $x = d$.

These conclusions are justified by the *Extreme Value Theorem* (that's why we need to assume continuity!).

Case 1: Either c or d is not an end-point, a or b. We now use the *Fermat's Theorem*: Every global extreme point has 0 derivative when it's not an end-point. It follows that $f'(c) = 0$ or $f'(d) = 0$, and we are done.

Case 2: Both c and d are end-points, a or b. Then

$$\left.\begin{array}{llll} f(c) & = f(a) & \text{or} & f(c) = f(b) \\ f(d) & = f(a) & \text{or} & f(d) = f(b) \end{array}\right\} \qquad \text{But these are equal!}$$

Therefore,

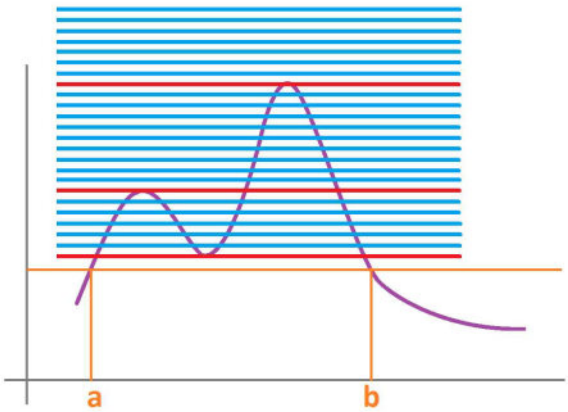
$$f(c) = f(d) \,.$$

But this means that

$$\max f = \min !$$

Therefore f must be constant on $[a, b]$. Therefore, $f'(x) = 0$ for all x in (a, b).

So, among all horizontal lines, there is at least one that touches the graph:



The theorem says that if you passed the same location twice, you must have stopped at some moment during this time. The former condition,

$$f(a) = f(b),$$

is rewritten as follows:

$$f(b) - f(a) = 0.$$

In other words, the *difference* is zero

$$\Delta f = 0.$$

It is the difference of f over the partition of $[a, b]$ with $n = 1$, $x_0 = a$, $x_1 = b$. Here are other ways to interpret this formula:

1. The change of the function over the interval is zero.
2. The rise over the interval is zero.
3. The displacement over the time interval is zero.

The condition can, furthermore, be rewritten as follows:

$$\frac{f(b) - f(a)}{b - a} = 0.$$

In other words, the *difference quotient*, the slope, is zero too:

$$\frac{\Delta f}{\Delta x} = 0.$$

Here are other ways to interpret this formula:

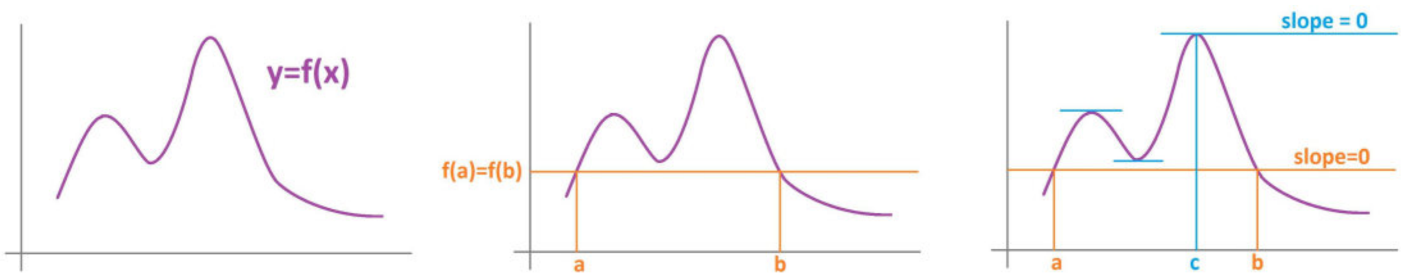
1. The average rate of change over the interval is zero.
2. The slope of the secant line is zero.
3. The average velocity over the time interval is zero.

So, the restrictions of the theorem are very limiting: It must be a *zero* velocity! Can it be 100 m/h? Let's investigate.

Rolle's Theorem says:

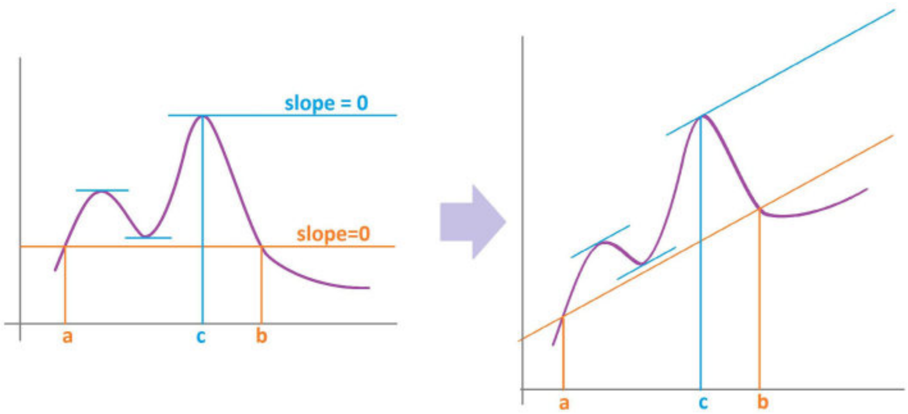
1. The average rate of change is zero $\Longrightarrow$ the instantaneous rate of change is zero at some point.
2. The slope of the secant line is zero $\Longrightarrow$ the slope of the tangent line is zero at some point.
3. The average velocity is zero $\Longrightarrow$ the instantaneous velocity is zero at some point.

The construction of the horizontal secant and tangent lines is outlined below:



What about non-horizontal lines that also *touch* the graph?

What happens if we *skew* the graph above? It's simple image editing:



The picture suggests what happens to the entities participating in Rolle's theorem:

- The tangents of those special points that used to be horizontal have become inclined.
- The secant line that connects the end-points that used to be horizontal has become inclined.

But these lines have remained *parallel*!

So, we need to link these two:

- the slopes of the tangent lines (that's the velocity) and
- the slope of the line between $(a, f(a))$ and $(b, f(b))$ (that's the average velocity).

Example 5.3.5: speeding

To illustrate the idea, let's revisit the issue of *speeding*.

A radar gun catches the *instantaneous* speed of your car. By law, it shouldn't be above 70 m/h. When the radar gun shows anything above, you're caught.

Now, suppose there is no radar gun. Imagine instead that a policeman observed you driving by (at a legal speed) and then, after 1 hour, another policeman observed you driving by – but 100 miles away!

Then the *average* speed of your car was 100 m/h. Did you violate the law? Considering that no-one can testify to have seen you drive above the speed limit, can the two policemen compare notes and prove that you did?

The analysis we pursue here allows them to *infer* that, yes, your instantaneous velocity was 100 m/h at some point. To make their case against you rock-solid, they'd quote the theorem below.

The following is a central result of calculus:

Theorem 5.3.6: Mean Value Theorem

Suppose we have a function:

1. *f is continuous on $[a, b]$.*
2. *f is differentiable on (a, b).*

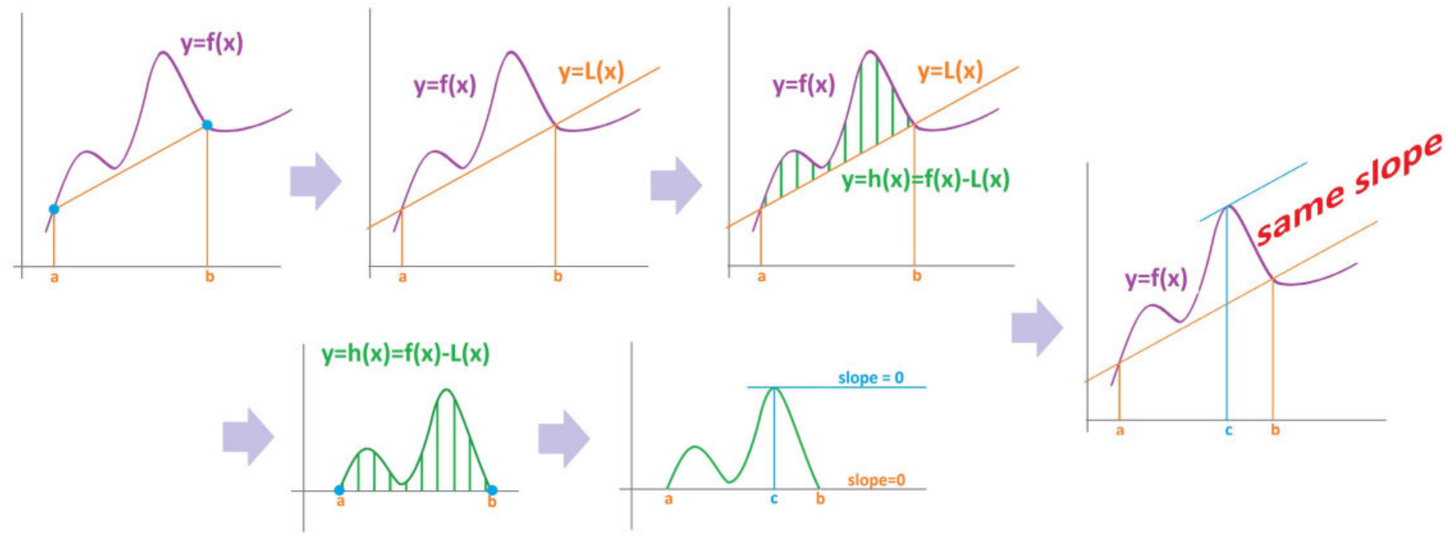
Then

$$\boxed{\frac{f(b) - f(a)}{b - a} = f'(c)}$$

FOR SOME c *in* (a, b) *.*

What happens if we have $f(a) = f(b)$ here? Then the left-hand side is 0, hence $0 = f'(c)$. We have the conclusion of Rolle's Theorem. This means that *MVT is more general than RT*. In other words, the latter is an instance, a narrow case of MVT.

The proof of MVT, however, will rely on RT. The idea is to "skew" the graph of MVT back to RT. This is the outline of the proof:



Proof.

Let's rename f in *Rolle's Theorem* as h to use it later. Then its conditions take this form:

1. h continuous on $[a, b]$.
2. h is differentiable on $[a, b]$.
3. $h(a) = h(b)$.

Suppose $y = L(x)$ is the linear function represented by the line between $(a, f(a))$ and $(b, f(b))$. Then, its derivative is simply the slope of the line:

$$L'(x) = \frac{f(b) - f(a)}{b - a}\,.$$

Now back to f. This is the key step; let

$$h(x) = f(x) - L(x)\,.$$

Let's verify the conditions above.

First, h is continuous on $[a, b]$ as the difference of the two continuous functions (SR). Condition #1 above is satisfied!

Next, h is differentiable on (a, b) as the difference of the two differentiable functions (SR). Condition

#2 above is satisfied! The derivative is simple:

$$h'(x) = f'(x) - \frac{f(b) - f(a)}{b - a}.$$

We also have:

$$f(a) = L(a),\ f(b) = L(b) \implies h(a) = 0,\ h(b) = 0 \implies h(a) = h(b).$$

Condition #3 above is satisfied!

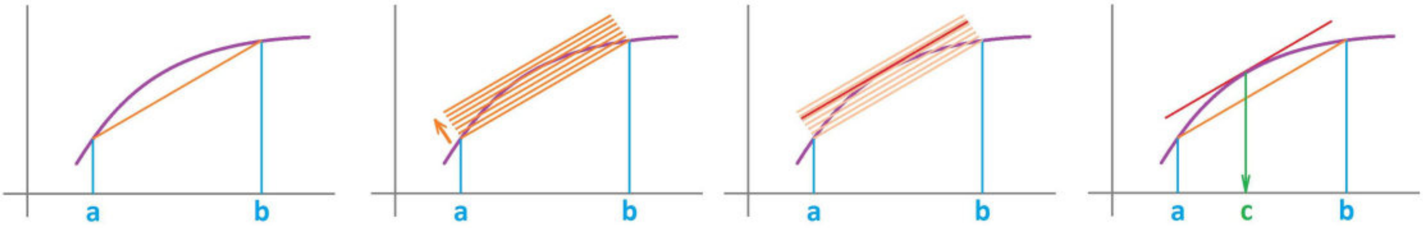
Thus, h satisfies the conditions of RT. Therefore, the conclusion is satisfied too:

$$h'(c) = 0$$

for some c in (a, b). In other words, we have:

$$f'(c) - \frac{f(b) - f(a)}{b - a} = 0.$$

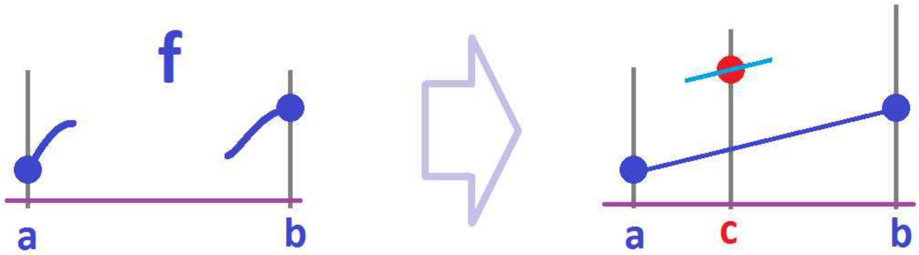
Geometrically, c is found by shifting the secant line until it touches the graph:



Exercise 5.3.7

Plot your own graph and find these special tangents.

We will use the Mean Value Theorem to derive facts about the values of a function we know nothing about from *a priori* information about its derivative:



Example 5.3.8: constant function

Let's try one of the converses mentioned in earlier in this section: The derivative of a constant function is zero, but are the constants the only functions with this property? Yes.

Suppose $f' = 0$ on $[A, B]$. Then for any $a < b$ in the interval, we have from the Mean Value Theorem:

$$\frac{f(b) - f(a)}{b - a} = f'(c) = 0.$$

Then,

$$f(b) = f(a).$$

The function is constant!

5.4. Monotonicity and the sign of the derivative

What does *a priori* information about the derivative of a function tell us about its monotonicity?

Recall that the Local Monotonicity Theorem states that, for a function $y = f(x)$ differentiable at $x = c$, if $x = c$ is an increasing point of $y = f(x)$, then $f'(c) \geq 0$. Instead of proving the converse of this *local* result, we will use the Mean Value Theorem to prove a *global* result!

We include the discrete case for completeness:

Theorem 5.4.1: Monotonicity from Sign of Difference Quotient

Suppose f is defined at the nodes of a partition on a closed interval I. Then:

1. *The difference quotient of f is non-negative on I if and only if f is increasing on I.*
2. *The difference quotient of f is non-positive on I if and only if f is decreasing on I.*

In other words, we have:

1. $\dfrac{\Delta f}{\Delta x} \geq 0 \iff f \nearrow$
2. $\dfrac{\Delta f}{\Delta x} \leq 0 \iff f \searrow$

The theorem for the derivatives below seems almost identical but it doesn't just follow with $\Delta x \to 0$ as many times before. We'll need to use the Mean Value Theorem to prove what we want:

Theorem 5.4.2: Monotonicity from Sign of Derivative

Suppose f is differentiable on an open interval I. Then:

1. *The derivative of f is non-negative on I if and only if f is increasing on I.*
2. *The derivative of f is non-positive on I if and only if f is decreasing on I.*

In other words, we have:

1. $\dfrac{df}{dx} \geq 0 \iff f \nearrow$
2. $\dfrac{df}{dx} \leq 0 \iff f \searrow$

Proof.

($\Leftarrow$) If f is increasing on I, every point c in I is an increasing point of f. Therefore, by the *Local Monotonicity Theorem*, we have $f'(c) \geq 0$.

($\Rightarrow$) Suppose a, b are in I and $a < b$. We need to show that $f(a) \leq f(b)$. By the *Mean Value Theorem*, we have:

$$\frac{f(b) - f(a)}{b - a} = f'(c)$$

for some c in (a, b). No matter what c is, by assumption of the theorem, the right-hand side is

non-negative. Therefore, we have

$$\frac{f(b)-f(a)}{b-a} \geq 0\,.$$

Now observe that in this fraction, the denominator is $b-a>0$. Therefore, the numerator must be also positive: $f(b)-f(a)\geq 0$. Hence, $f(b)\geq f(a)$.

Example 5.4.3: three familiar functions

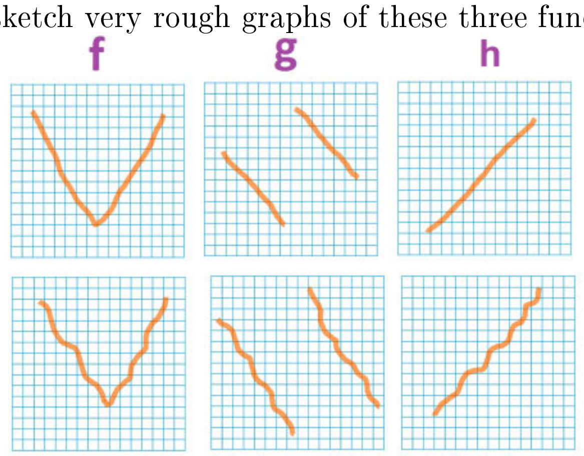
We use the derivatives to analyze these functions:

$$\begin{array}{lllll}
(1) & f(x)=3x^2+1 & \Longrightarrow & f'(x)=6x+3 & \Longrightarrow \\
 & f'(x)<0 \text{ if } x<-1/2 & \text{and} & f'(x)>0 \text{ if } x>-1/2 & \Longrightarrow \\
 & f \searrow \text{ on } (-\infty,-1/2) & \text{and} & f \nearrow \text{ on } (-1/2,\infty) &
\end{array}$$

$$\begin{array}{lllll}
(2) & g(x)=\dfrac{1}{x} & \Longrightarrow & g'(x)=-\dfrac{1}{x^2} & \Longrightarrow \\
 & g'(x)<0 \text{ if } x<0 & \text{and} & g'(x)<0 \text{ if } x>0 & \Longrightarrow \\
 & g \searrow \text{ on } (-\infty,0) & \text{and} & g \searrow \text{ on } (0,\infty) &
\end{array}$$

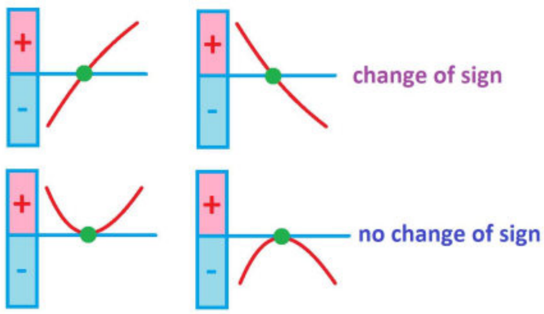
$$\begin{array}{lllll}
(3) & h(x)=e^x & \Longrightarrow & h'(x)=e^x & \Longrightarrow \\
 & h'(x)>0 \text{ for all } x & \Longrightarrow & & \\
 & h \nearrow \text{ on } (-\infty,\infty) & & &
\end{array}$$

With just this data, we can sketch very rough graphs of these three functions:



Even though we have shown that the functions have no other extrema, the curves could have these "wiggles". This issue will be addressed shortly.

To identify monotonicity, therefore, will require a *sign analysis* (seen in Volume 1, Chapter 1PC-4) of the derivative, as follows:



The main tool is the *Intermediate Value Theorem* (Chapter 2):

▶ If a function is continuous, its sign can only change where it is zero.

Example 5.4.4: quadratic

Recall this example:

$$f(x) = x^3 - 3x\,.$$

On what intervals is this function increasing and decreasing?

First,

$$f'(x) = 3x^2 - 3\,.$$

In order to use the *Monotonicity Theorem*, we need to find those x's that produce $f'(x) > 0$ or $f'(x) < 0$. In other words, we need to solve those *inequalities*.

We start with the corresponding *equation* $f'(x) = 0$, done previously:

$$\begin{array}{lll} 3x^2 - 3 & = 0 & \Longrightarrow \\ x^2 - 1 & = 0 & \Longrightarrow \\ x^2 & = 1 & \Longrightarrow \\ x & = \pm 1\,. & \end{array}$$

These two points cut three intervals from the real line:

- $(-\infty, -1)$,
- $(-1, 1)$,
- $(1, \infty)$.

We need to know the *sign of the derivative* on each.

Since f' is continuous, the sign of f' can only change at -1 or 1. Therefore, we just need to *sample one point* within each interval to determine the sign of the derivative on the whole interval:

- Pick $x = -2$, then $f'(-2) = 3 \cdot (-2)^2 - 3 = 9 > 0$. Therefore, $f'(x) > 0$ for x in $(-\infty, -1)$. Therefore, $f \nearrow$ on $(-\infty, -1)$.
- Pick $x = 0$, then $f'(0) = -3 < 0$. Therefore, $f'(x) < 0$ for x in $(-1, 1)$. Therefore, $f \searrow$ on $(-1, 1)$.
- Pick $x = 2$, then $f'(2) = 3 \cdot 2^2 - 3 = 9 > 0$. Therefore, $f'(x) > 0$ for x in $(1, \infty)$. Therefore, $f \nearrow$ on $(1, \infty)$.

Let's put this data in a table:

$$\begin{array}{lccc} x: & (-\infty, -1) & (-1, 1) & (1, \infty) \\ f: & \nearrow & \searrow & \nearrow \end{array}$$

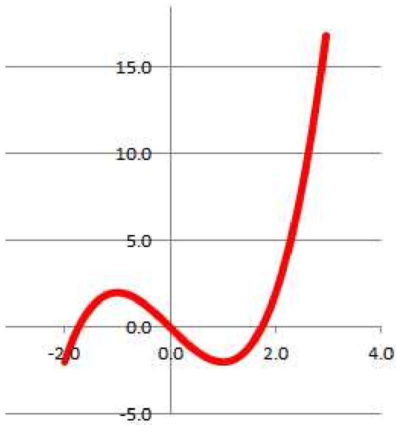
This is the answer!

Furthermore, these arrows are close to forming a curve, especially after this modification:

$$\begin{array}{lccccc}
 & & \cdot & & & \\
f: & \nearrow & & \searrow & & \nearrow \\
 & & & & \cdot & \\
x: & & -1 & & 1 & \\
 & & \text{max} & & \text{min} &
\end{array}$$

We have also – automatically – classified the extreme points!

We confirm the result by plotting:



Example 5.4.5: sine

Consider $f(x) = \sin x$ again. We already know that the difference quotient and the derivative are zero, $\cos x = 0$, at these locations:

$$x = \frac{\pi}{2} + k\pi,\ \ k = 0, \pm 1, \pm 2, ...$$

These are the points where the difference quotient and the derivative could, potentially, change their sign! Since the two differ by a positive multiple:

$$\frac{\Delta}{\Delta x}(\cos x) = -\frac{\sin(h/2)}{h/2} \cdot \sin x = \frac{\sin(h/2)}{h/2} \cdot \frac{d}{dx}(\cos x)\,,$$

they change signs together. From what we know even more about cosine, the sign does change every time:

$$\begin{array}{rccccccccccccc}
 & & & \cdot & & & & \cdot & & & & \cdot & & \\
y = \sin x: & & \nearrow & & \searrow & & \nearrow & & \searrow & & \nearrow & & & \\
 & \cdot & & & & \cdot & & & & \cdot & & & & \\
x: & -\pi/2 & & \pi/2 & & 3\pi/2 & & 5\pi/2 & & 7\pi/2 & & 9\pi/2 & & \\
y = \cos x: & 0 & + & 0 & - & 0 & + & 0 & - & 0 & + & 0 & - &
\end{array}$$

This conclusion confirms what we know about this function:

Example 5.4.6: exponent

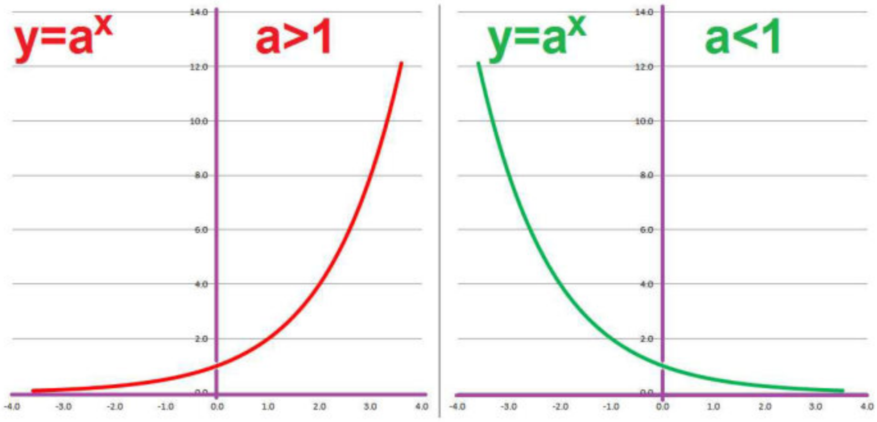
Let's consider the exponential function $f(x) = a^x$ for all $a > 0$. First the derivatives:

$$f'(x) = (a^x)' = a^x \ln a\,.$$

Therefore, by the *Monotonicity Theorem*, for all x we have:

$$\begin{array}{lll} a > 1 & \Longrightarrow & f'(x) > 0 \\ 0 < a < 1 & \Longrightarrow & f'(x) < 0 \end{array}$$

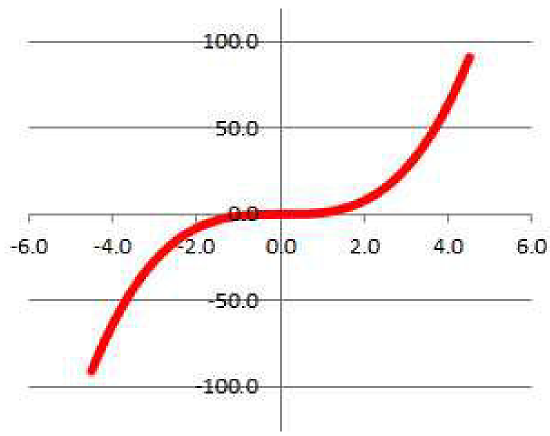
It follows that the function is either all increasing or all decreasing:



Then, when $a > 1$, we have the global minimum on interval $[a, b]$ at $x = a$ and the global maximum at b.

Example 5.4.7: cube

Let's consider $f(x) = x^3$.



The derivative is $f'(x) = 3x^2$. Therefore, by the *Monotonicity Theorem*, we have:

$$\begin{array}{llll} \text{on } (-\infty, 0) & f' > 0 & \Longrightarrow & f \nearrow \\ \text{at } 0 & f' = 0 & & \\ \text{on } (0, +\infty) & f' > 0 & \Longrightarrow & f \nearrow \end{array}$$

There is more here: We have found an example of a strictly increasing function with a zero derivative at one point!

These are a modified version of the above theorems:

Theorem 5.4.8: Strict Monotonicity via Difference Quotient

Suppose f is defined at the nodes of a partition on a closed interval I. Then, we have:

1. *The difference quotient of f is positive on I if and only if f is strictly*

increasing on I.

2. *The difference quotient of f is negative on I if and only if f is strictly decreasing on I.*

In other words, we have:

1. $\frac{\Delta f}{\Delta x} > 0 \iff f \nearrow$ *strictly*
2. $\frac{\Delta f}{\Delta x} < 0 \iff f \searrow$ *strictly*

The continuous case is more subtle:

Theorem 5.4.9: Strict Monotonicity via Derivative

Suppose f is differentiable on an open interval I. Then, we have:

1. *If the derivative of f is positive on I, then f is strictly increasing on I.*
2. *If the derivative of f is negative on I, then f is strictly decreasing on I.*

In other words, we have:

1. $\frac{df}{dx} > 0 \implies f \nearrow$ *strictly*
2. $\frac{df}{dx} < 0 \implies f \searrow$ *strictly*

Proof.

Just replace each "$\geq$" in the proof of the *Monotonicity Theorem* with "$>$".

We demonstrated in the last example that the converse fails. The reason for such a difference from the non-strict case is, roughly, that the limit of positive numbers might be zero.

Exercise 5.4.10

Show that if the derivative of a function is zero, it is a constant function, in two ways: (a) by modifying the proof of the Monotonicity Theorem, (b) by applying the Monotonicity Theorem.

A bonus result is below:

Corollary 5.4.11: One-to-one from Derivative

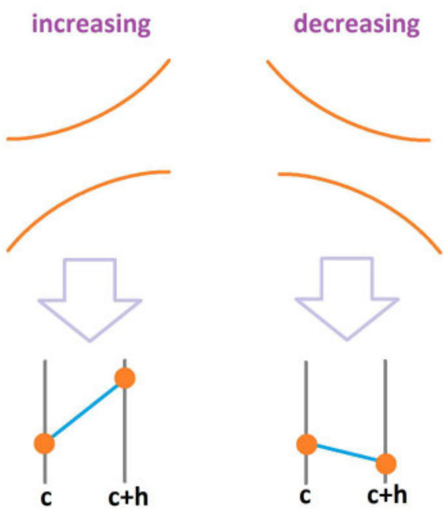
Suppose f is differentiable on an open interval I. Then, if $f' > 0$ on I or $f' < 0$ on I, then f is one-to-one on I.

Exercise 5.4.12

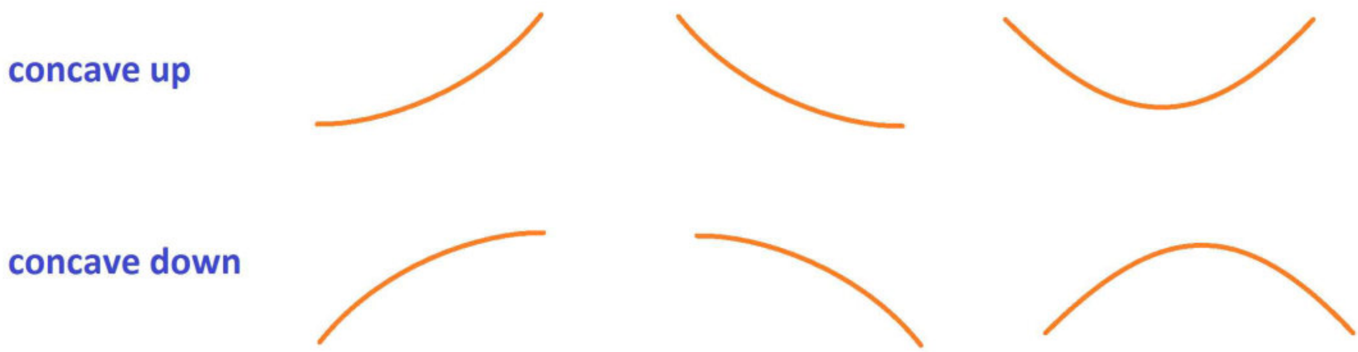
Prove the corollary. What about the converse?

5.5. Concavity and the sign of the second derivative

Recall how easy it is to see the idea behind monotonicity by zooming in on the graph when it is made of overlapping dots:

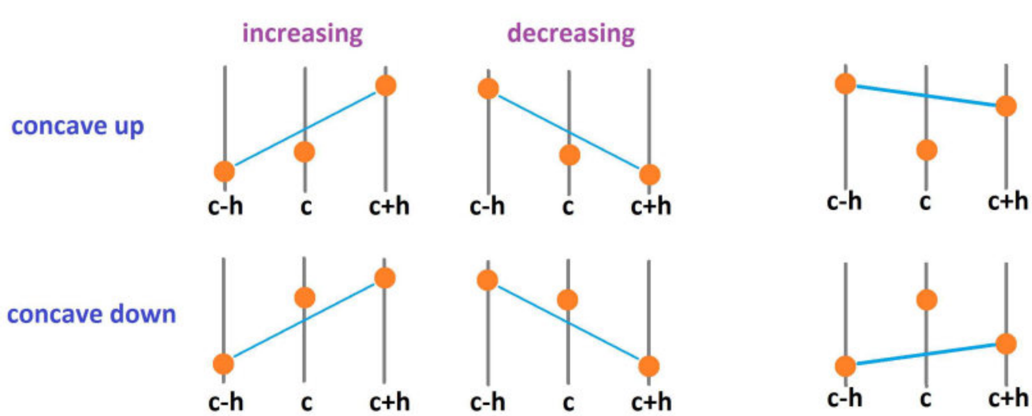


We will do the same for a subtler idea. The picture below informally reveals the meaning of upward and downward *concavity*:



When we guarantee the one or the other, we eliminate the possibility of "wiggly" curves! We will also be able to tell maxima from minima.

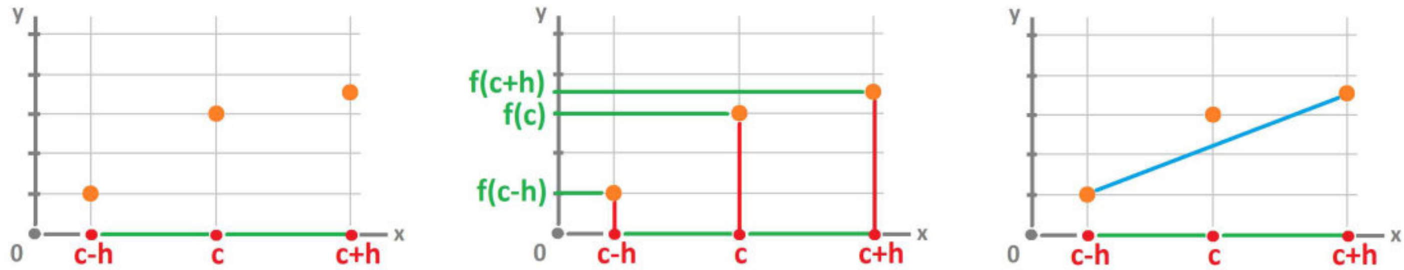
Let's zoom in. There are *three* points this time:

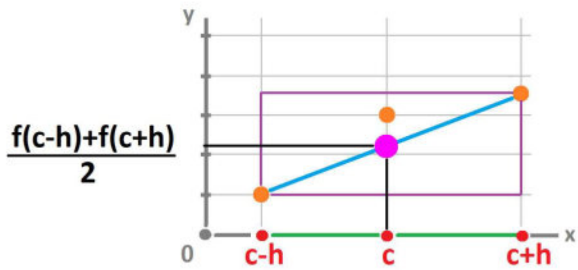


The pattern is clear:

> ▶ It is concave up when the middle point lies below the line connecting the other two, and concave down when it is above.

Let's investigate the algebra. The two points on the y-axis are $f(c-h)$ and $f(c+h)$:





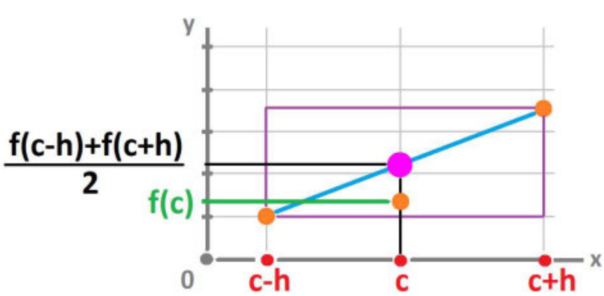
The y-value of the mid-point is their average:

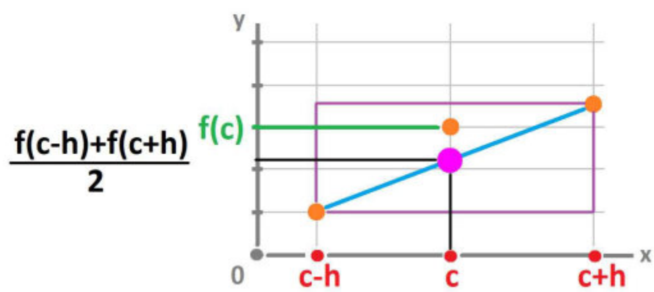
$$\frac{f(c-h)+f(c+h)}{2}\,.$$

Then, the concave up and concave down conditions take the form of these inequalities respectively:

$$f(c)\le\frac{f(c-h)+f(c+h)}{2}\qquad\text{and}\qquad f(c)\ge\frac{f(c-h)+f(c+h)}{2}\,.$$

They are illustrated below:





What does this expression have to do with the differences or the derivatives?

Let's re-arrange the terms in the first inequality:

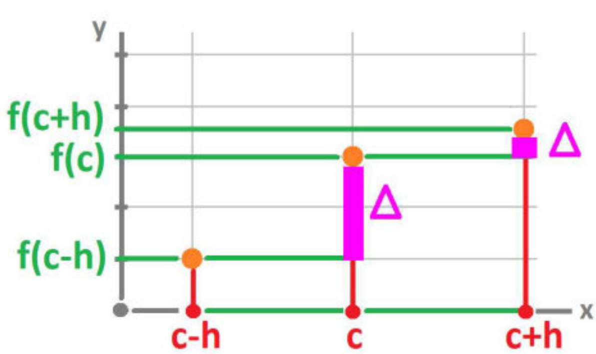
$$f(c+h)-2f(c)+f(c-h)\ge 0\,.$$

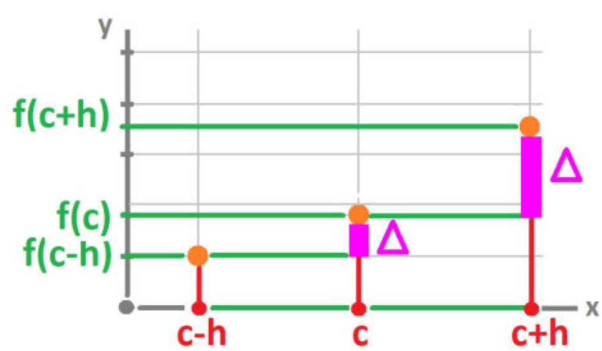
A bit more and we see two *differences*:

$$\big[f(c+h)-f(c)\big]-\big[f(c)-f(c-h)\big]\ge 0\,.$$

In fact, this is the *difference of the differences* (or the change of the change etc.)!

When the second difference is smaller than the first, we have it concave down; otherwise, up:





If we assume the mid-point secondary nodes, the concavity condition above becomes:

$$\Delta f(c+h/2) - \Delta f(c-h/2) \ge 0\,.$$

It only takes division by $h>0$, twice, to arrive at *the difference quotient of the difference quotient* at $x=c$:

$$\frac{\dfrac{\Delta f(c+h/2)}{h} - \dfrac{\Delta f(c-h/2)}{h}}{h} \ge 0\,,$$

This is just the *second difference quotient* (Chapter 4):

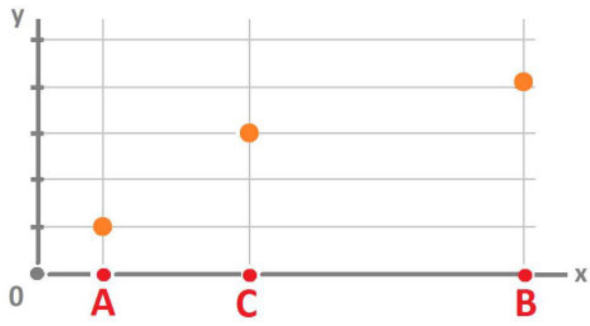
$$\boxed{\frac{\Delta^2 f}{\Delta x^2}(c) \ge 0}$$

So, there is a match:

- We know that the *sign* of the difference quotient or the derivative tells us the difference between increasing and decreasing behavior.
- We have discovered that the *sign* of the second difference quotient or the second derivative tells us the difference between the upward and downward concavity.

We need a more general point of view: What if the points aren't equally spaced?

Suppose we have three consecutive points A, B, C on the x-axis:



We still need to express C in terms of A and B:

$$C = \alpha A + \beta B\,,$$

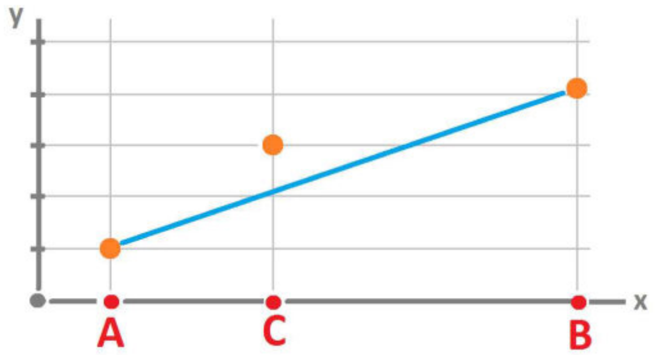
for some pair of numbers $\alpha, \beta \ge 0$ with $\alpha + \beta = 1$. In other words, C is a *weighted average* of A and B. The numbers α and β are the "weights":

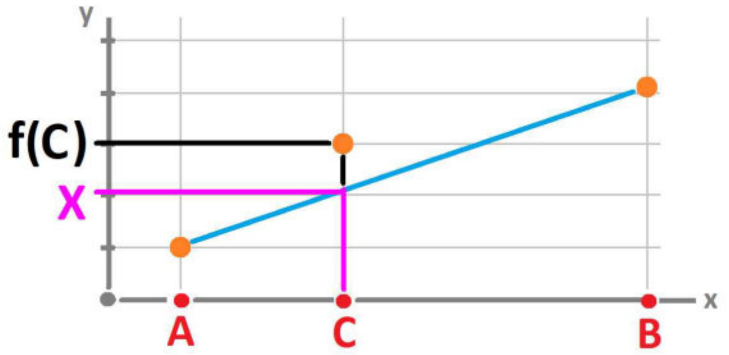
To look at this differently, these two numbers α and β give us the relative position of C within the interval $[A, B]$.

Now the y's. Consider the point:

$$X = \alpha f(A) + \beta f(B)\,.$$

Once again, this is a *weighted average* of $f(A)$ and $f(B)$. The numbers α and β also give us the relative position of X within the interval formed by $f(A)$ and $f(B)$:





Exercise 5.5.1

Show that for a linear function, the weights remain the same:

$$f(\alpha A+\beta B)=\alpha f(A)+\beta f(B)\,.$$

What can you say about the converse?

The concavity is determined by whether $f(C)$ is below or above X, as follows:

Definition 5.5.2: concave on partition

Suppose three consecutive nodes A,B,C of a partition satisfy

$$C=\alpha A+\beta B\,,$$

with $\alpha\ge 0$ and $\beta\ge 0$ and $\alpha+\beta=1$. For a function f defined on these nodes, we define:

1. The function f is called *concave up* at B when
$$f(C)\le \alpha f(A)+\beta f(B)\,.$$
2. The function f is called *concave down* at B when
$$f(C)\ge \alpha f(A)+\beta f(B)\,.$$

When $\alpha=\beta=1/2$, the definition produces the inequalities in the analysis we presented in the beginning of the section.

We will occasionally use the notation:

Concavity

- $\smile$ concave up
- $\frown$ concave down

Above we proved the following for the case of equally distributed nodes:

Theorem 5.5.3: Concavity on Partition

Suppose a function f is defined at the nodes of a partition of interval $[a,b]$. Then we have:

1. *The function f is concave up on $[a,b]$ if and only if the second difference quotient is non-negative.*
2. *The function f is concave down on $[a,b]$ if and only if the second difference quotient is non-positive.*

In other words, we have:

1. $f \smile \iff \dfrac{\Delta^2 f}{\Delta x^2} \geq 0$
2. $f \frown \iff \dfrac{\Delta^2 f}{\Delta x^2} \leq 0$

Proof.

Suppose three consecutive nodes x_{i-1}, x_i, x_{i+1} of a partition satisfy

$$x_i = \alpha x_{i-1} + \beta x_{i+1}\,,$$

and

$$f(x_i) \leq \alpha f(x_{i-1}) + \beta f(x_{i+1})\,,$$

for some pair $\alpha \geq 0$ and $\beta \geq 0$ with $\alpha + \beta = 1$.

Let's first find α and β. Consider the first equation of the definition rewritten:

$$x_i = \alpha(x_i - \Delta x_{i-1}) + \beta(x_i + \Delta x_i)\,.$$

Cancellation produces the following:

$$0 = \alpha(-\Delta x_{i-1}) + \beta(\Delta x_i)\,.$$

Therefore,

$$\alpha = \frac{\Delta x_i}{\Delta x_{i-1} + \Delta x_i},\ \beta = \frac{\Delta x_{i-1}}{\Delta x_{i-1} + \Delta x_i}\,.$$

The concavity is determined by the *sign* of the following expression:

$$\alpha f(x_{i-1}) + \beta f(x_{i+1}) - f(x_i) \geq 0\,.$$

What is its meaning? We substitute α and β and our expression becomes:

$$\frac{\Delta x_i}{\Delta x_{i-1} + \Delta x_i} f(x_{i-1}) + \frac{\Delta x_{i-1}}{\Delta x_{i-1} + \Delta x_i} f(x_{i+1}) - f(x_i) \geq 0\,.$$

Let's rearrange the terms:

$$\Delta x_i f(x_{i-1}) + \Delta x_{i-1} f(x_{i+1}) - (\Delta x_{i-1} + \Delta x_i) f(x_i) \geq 0\,,$$

and factor:

$$\Delta x_{i-1}\big(f(x_{i+1}) - f(x_i)\big) - \Delta x_k\big(f(x_i) - f(x_{i-1})\big) \geq 0\,.$$

In parentheses we see the two differences of f evaluated at the two adjacent intervals $[x_{i-1}, x_i], [x_i, x_{i+1}]$. To see the difference *quotients*, let's divide this by $\Delta x_{i-1}\Delta x_i$:

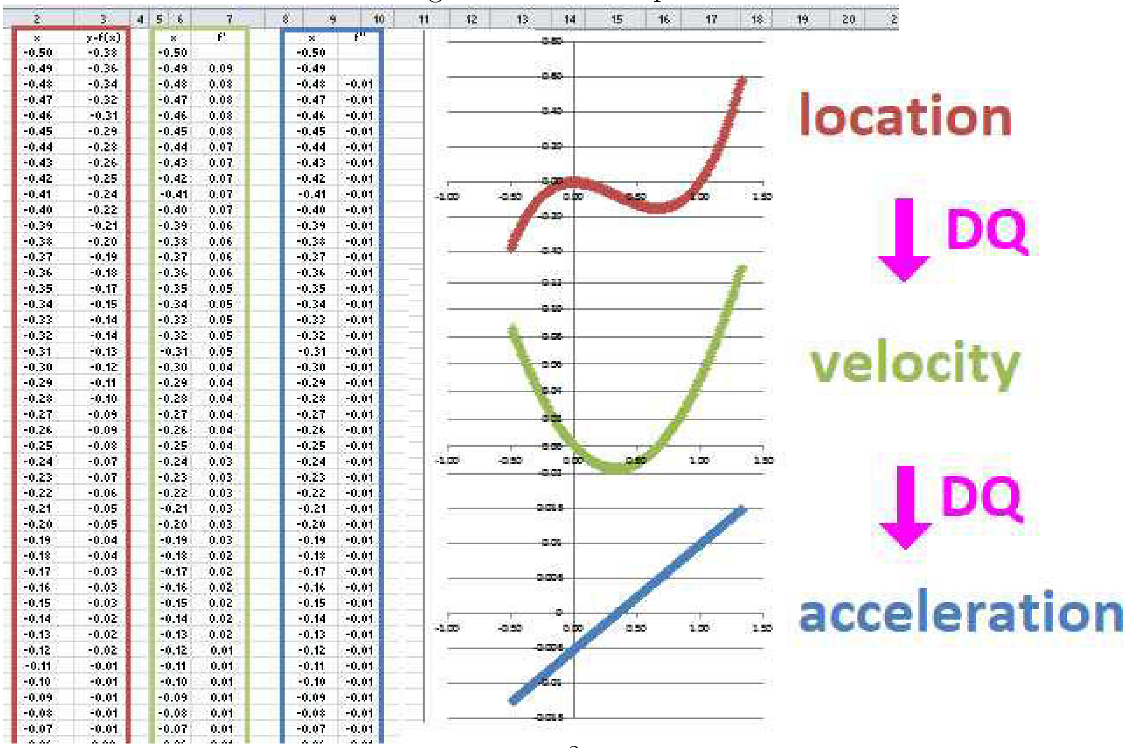
$$\frac{f(x_{i+1}) - f(x_i)}{\Delta x_i} - \frac{f(x_i) - f(x_{i-1})}{\Delta x_{i-1}} \geq 0 \quad \text{or} \quad \frac{\Delta f}{\Delta x}(c_i) - \frac{\Delta f}{\Delta x}(c_{i-1}) \geq 0\,.$$

It only takes division by Δc_i to arrive at the difference quotient of the difference quotient:

$$\frac{\Delta^2 f}{\Delta x^2}(x_i) \geq 0\,.$$

Example 5.5.4: spreadsheet

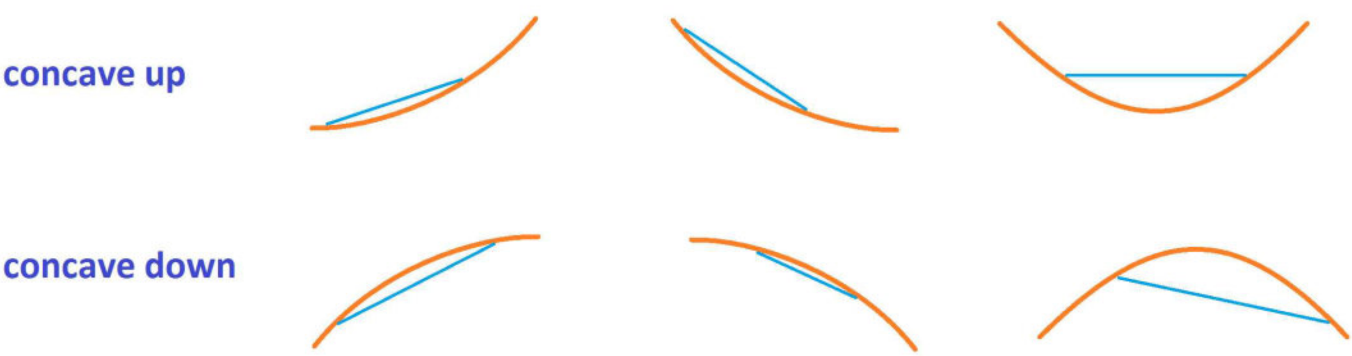
The second difference quotient can be computed with a spreadsheet. Recall (Chapter 4) that it also represents the acceleration when the original function represents the location:



The transitions from f to $\dfrac{\Delta f}{\Delta x}$ and from $\dfrac{\Delta f}{\Delta x}$ to $\dfrac{\Delta^2 f}{\Delta x^2}$ are implemented with the same formula.

Now, to the *continuous case*:

A function is concave when any of its samplings is. In other words, the segments of the secant lines ("cords") lie either above or below the graph:



All of the cords, not just the ones that connect the ends! We have to carefully phrase these requirements:

Definition 5.5.5: concavity on interval

1. If for each x in I, each $k, h > 0$, and any pair of positive α and β with $\alpha + \beta = 1$, we have:
$$f(x) \leq \alpha f(x-k) + \beta f(x+h)\,,$$
within the interval I, then the function f is called *concave up* on interval I.
2. When the opposite inequality is satisfied:
$$f(x) \geq \alpha f(x-k) + \beta f(x+h)\,,$$
the function f is called *concave down* on interval I.

The function is *strictly concave* (up or down) if the inequality is strict.

Exercise 5.5.6

Is a linear function concave up or down?

The derivative makes complex things look simple:

Theorem 5.5.7: Concavity Theorem

Suppose a function f is twice differentiable on an open interval I. Then, we have:

1. *If $f'' \geq 0$ on I, then f is concave up on I.*
2. *If $f'' \leq 0$ on I, then f concave down on I.*

Proof.

If $f'' \geq 0$ on I, then by the *Monotonicity Theorem*, f' is increasing. Therefore,

$$f'(s) \leq f'(t)$$

for all $s < t$ in I. Now by the *Mean Value Theorem*, we have:

$$\frac{f(c) - f(c-h)}{h} = f'(s)$$

for some s in $(c-h, c)$, and

$$\frac{f(c+h) - f(c)}{h} = f'(t)$$

for some t in $(c, c+h)$. Therefore,

$$\big(f(c-h) + f(c+h)\big) - 2f(c) = h\left(\frac{f(c+h) - f(c)}{h} - \frac{f(c) - f(c-h)}{h}\right) = h(f'(t) - f'(s)) \geq 0\,.$$

Exercise 5.5.8

Derive the theorem from the theorem about discrete concavity in this section.

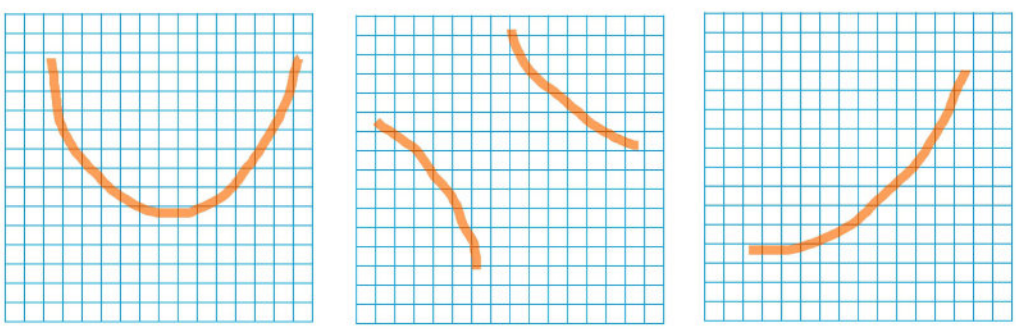
Example 5.5.9: three familiar functions

$$
\begin{aligned}
(1)\quad & f(x) && = 3x^2 + 1 \implies \\
& f'(x) && = 6x \implies \\
& f''(x) && = 6 \\
& f''(x) > 0 && \implies f \smile
\end{aligned}
$$

$$
\begin{aligned}
(2)\quad & g(x) && = \frac{1}{x} \implies \\
& g'(x) && = -\frac{1}{x^2} \text{ for } x \neq 0 \implies \\
& g''(x) && = \frac{1}{2x^3} \text{ for } x \neq 0 \\
& g''(x) < 0 \text{ for } x < 0 && \implies g \frown \text{ on } (-\infty, 0) \\
& g''(x) > 0 \text{ for } x > 0 && \implies g \smile \text{ on } (0, \infty)
\end{aligned}
$$

$$
\begin{aligned}
(3)\quad & h(x) && = e^x \implies \\
& h'(x) && = e^x \implies \\
& h''(x) && = e^x \\
& h''(x) > 0 && \implies h \smile
\end{aligned}
$$

Using this data combined with the increasing/decreasing behavior established earlier, we can sketch by hand the graphs of these three functions:

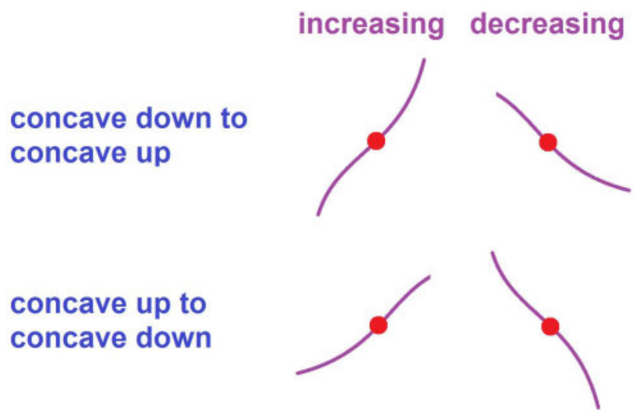


f **g** **h**

Exercise 5.5.10

Point out where the graphs are inadequate.

Now, the points where the behavior changes are especially important:



Definition 5.5.11: inflection point

The points where the function changes its concavity are called *inflection points*; i.e., these are such points c that the function's concavity on some open interval (a,c) is opposite of the concavity on some open interval (c,b).

Warning!

The value of the first derivative at inflection points can be arbitrary; it is 0 for $f(x) = x^3$ at 0 and non-zero for the trigonometric functions below.

Example 5.5.12: sine and cosine

If we use the Trig Formulas:

$$\begin{aligned}(\sin x)' &= \cos x,\\ (\cos x)' &= -\sin x.\end{aligned}$$

twice, we have:

$$\begin{aligned}(\sin x)'' &= -\sin x,\\ (\cos x)'' &= -\cos x.\end{aligned}$$

As we know, the two functions alternate – periodically every π – between positive and negative values:

$$+ \quad 0 \quad - \quad + \quad 0 \quad - \quad 0 \quad + \quad 0-$$

Therefore, the two functions alternate – periodically every π – between concave up and concave down behavior:

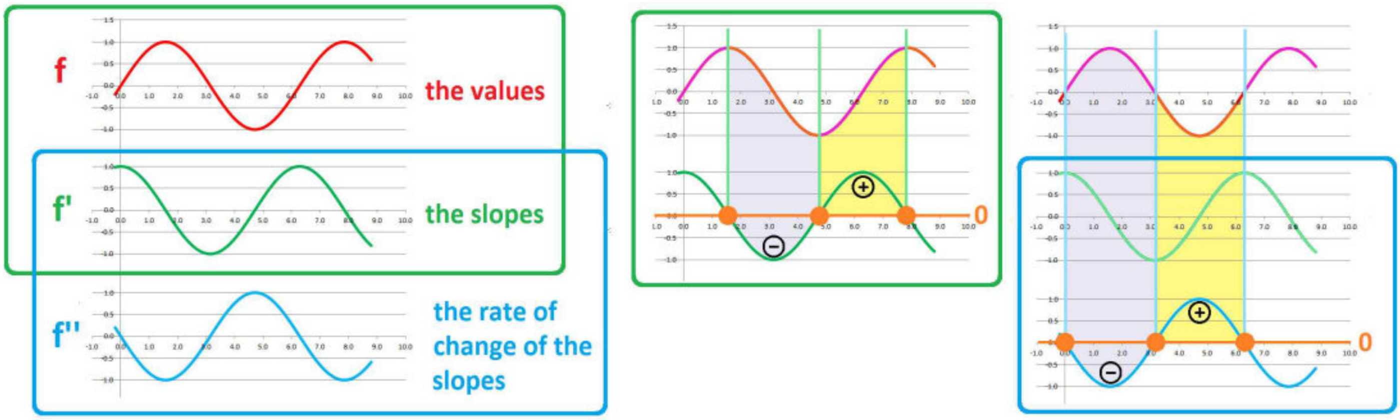
$$\nearrow \quad \frown \quad \searrow \quad \smile \quad \nearrow \quad \frown \quad \searrow \quad \smile \quad \nearrow \quad \frown \quad \searrow \quad \smile$$

These are the inflection points of the function.

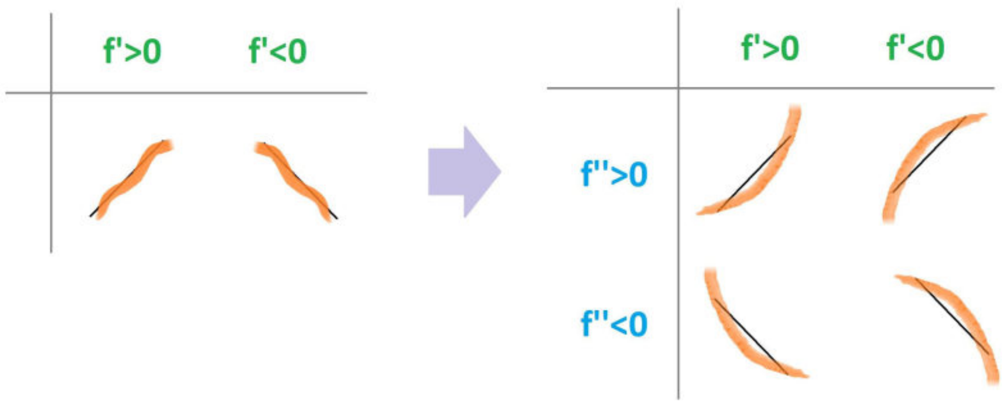
There are parallels with our activities earlier in this chapter:

- When using the *Monotonicity Theorem*, we compare the *shapes* of the patches of the graph of the function f, $\searrow$ or $\nearrow$, to the signs of the *values* of the first derivative f', $+$ or $-$.
- Meanwhile, when using the *Concavity Theorem*, we compare the *shapes* of the patches of the graph of the function f, $\smile$ or $\frown$, to the signs of the *values* of the second derivative f'', $+$ or $-$.

These outcomes are shown below:



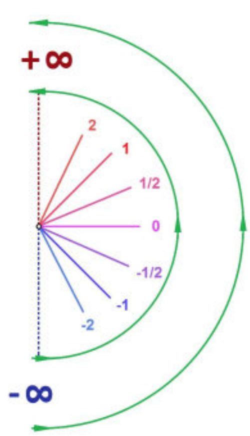
One can see how we use a higher level of analysis in comparison to the first derivative:



The wiggles are gone!

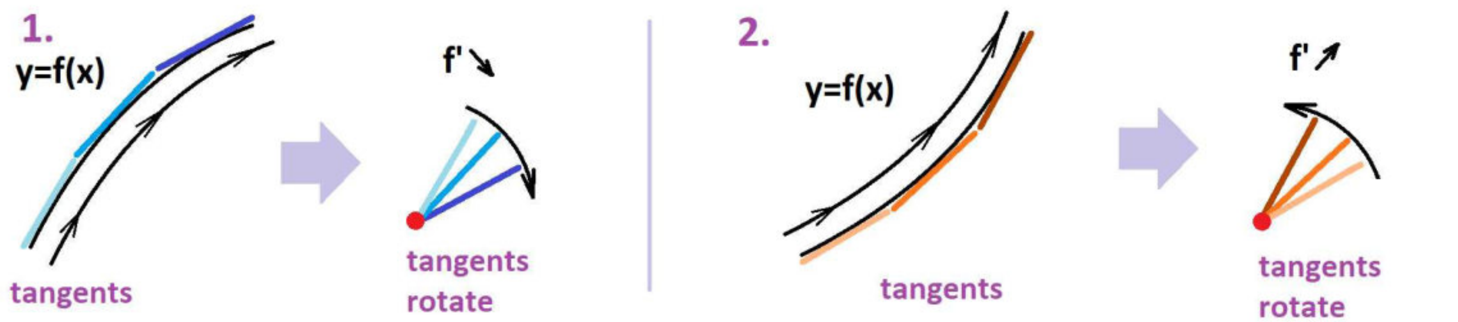
> **Warning!**
>
> As the second diagram indicate, monotonicity and concavity are two *independent* characteristics of a function; one should never try to figure out one from the other.

The conclusion of the *Concavity Theorem* is seen from the following: As the first derivative represents the slopes of the function, the second derivative represents the rate of change of these slopes. As you can see, the slopes increase when the lines are rotated in the counter-clockwise direction:



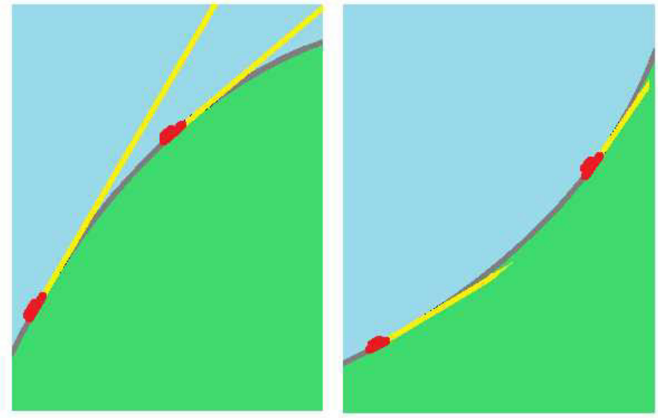
Therefore, the tangent lines rotate as follows:

- Decreasing slopes $\Longrightarrow$ tangent lines rotate clockwise.
- Increasing slopes $\Longrightarrow$ tangent lines rotate counter-clockwise.



Example 5.5.13: climbing

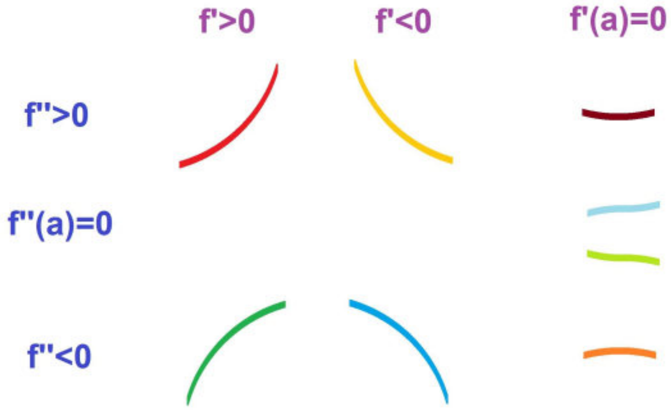
We can appreciate this idea if we imagine how we drive on this road:



Even though we are *climbing* in either case, what happens to the beams is different:

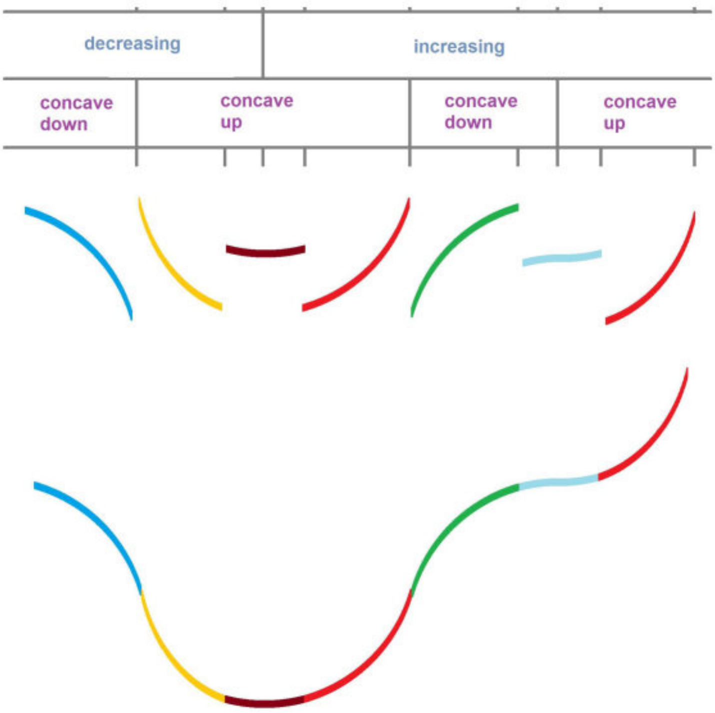
- When the road is concave *down*, the headlights point up above the road. The driver might anticipate less need for gas.
- When the road is concave *up*, the headlights point down into the road. The driver might anticipate more need for gas.

All graphs are made of the following eight pieces, classified according to the sign of the derivative and the second derivative:



Example 5.5.14: piecing together

Below, we start with the information about the monotonicity and concavity of a function on several intervals, then pick appropriate pieces from the above table, and then glue them together to form a continuous curve:



Here is yet another way to interpret the terminology and the notation:

"feeling up"	"feeling down"
$\ddot{\smile}$	$\ddot{\frown}$

5.6. Derivatives and extrema

Recall what the *Fermat's Theorem* does and does not say about a function differentiable at $x = c$:

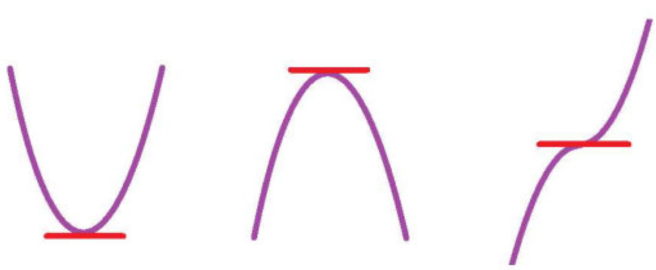
$$x = c \text{ is } \left| \begin{array}{l} \text{a local} \\ \text{max/min} \end{array} \right| \begin{array}{c} \Longrightarrow \\ \not\Longleftarrow \end{array} \quad \boxed{f'(c) = 0}$$

When *can* we reverse the arrow?

When $f'(c) = 0$, there are these three possibilities for c:

$$\text{max}, \ \text{min}, \ \ \text{neither}.$$

At their simplest, they look like this:



It is possible to tell one from another by looking at the derivative *in the vicinity* of the point.

Example 5.6.1: cubic

Consider again:

$$f(x) = x^3 - 3x\,.$$

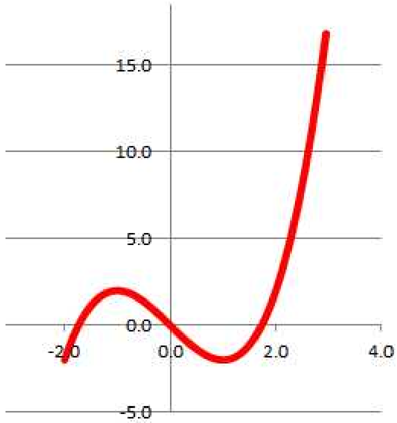
We know that $f'(x) = 3x^2 - 3 = 0$ for $x = \pm 1$ and none others. The derivative is 0, so these *may be* extreme points.

How do we find out? We look at the *signs of the derivative*:

- $f' > 0$ and $f \nearrow$ to the left of -1 .
- $f' < 0$ and $f \searrow$ to the right of -1.

This can only happen when $x = -1$ is a local max. The opposite for $x = 1$.

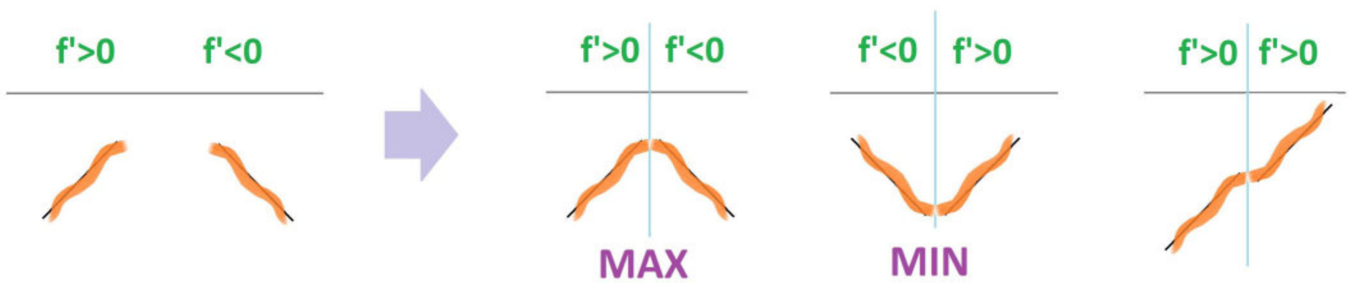
We confirm with a plot:



We have learned that we can classify critical points by looking at the *change of monotonicity*:

$\nearrow\searrow$ max $\quad$ $\searrow\nearrow$ min

Furthermore, the function's monotonicity is determined by the *sign* of its first derivative:



We see the *Monotonicity Theorem* on the left and our conclusion on the right.

The summary result is the following:

Theorem 5.6.2: First Derivative Test

If the derivative changes its sign at a point, the point is an extremum.

In other words, we have the following. Suppose f is differentiable on an open interval I that contains point $x = c$. Then, we have:

1. *If $f'(x) \geq 0$ for all $x < c$* `AND` *$f'(x) \leq 0$ for all $x > c$ within I, then c is a local maximum point.*
2. *If $f'(x) \leq 0$ for all $x < c$* `AND` *$f'(x) \geq 0$ for all $x > c$ within I, then c is a local minimum point.*

3. *If $f'(x) \le 0$ for all x in I* OR *$f'(x) \ge 0$ for all x in I, then c is neither a local maximum nor minimum point.*

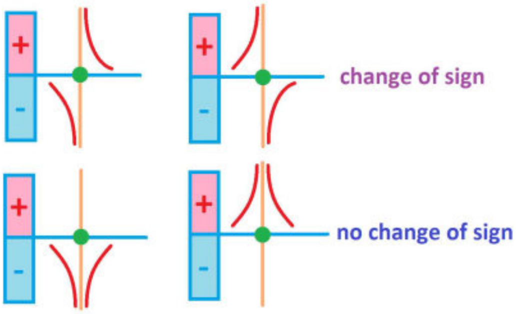
Proof.

Suppose $I = (a, b)$. If $f'(x) \ge 0$ for all $a < x < c$, then by the *Monotonicity Theorem*, $f(x) \le f(c)$ for all $a < x < c$. If $f'(x) \le 0$ for all $c < x < b$, then by the Monotonicity Theorem, $f(x) \le f(c)$ for all $c < x < b$. Thus, $f(x) \le f(c)$ for all $a < x < b$, $x \ne c$. Then, c is a local maximum.

In other words, we have:

- If $f'(x)$ changes its sign at $x = c$ from $+$ to $-$, then c is a local maximum point.
- If $f'(x)$ changes its sign at $x = c$ from $-$ to $+$, then c is a local minimum point.

To classify critical points, again, will require a *sign analysis* (seen in Volume 1, Chapter 1PC-4) of the derivative. It was done in the last example for a polynomial. In case of rational function, we have to take into account the possibility that a change of sign happens across a vertical asymptote:



.

Example 5.6.3: rational function

Let's analyze this function:

$$f(x) = \frac{x^3}{x^2 - 3}.$$

First, the domain is all x except $\pm\sqrt{3}$.

Next, we differentiate:

$$\begin{aligned} f'(x) &= \left(\frac{x^3}{x^2-3}\right)' \\ &= \frac{3x^2(x^2-3) - x^3 2x}{(x^2-3)^2} && \text{This will have to be subjected to sign analysis.} \\ &= \frac{x^4 - 9x^2}{(x^2-3)^2} && \text{We need to factor it!} \\ &= \frac{x^2(x-3)(x+3)}{(x-\sqrt{3})^2(x+\sqrt{3})^2}. \end{aligned}$$

Next, the critical points are the ones where one of these two things happens:

- The derivative is zero, i.e., the numerator is zero. We read these from its factors (as seen in Volume 1, Chapter 1PC-4):

$$x = 0,\ 3,\ -3.$$

- The derivative is undefined, i.e., the denominator is zero. We read these from its factors:

$$x = -\sqrt{3},\ \sqrt{3}.$$

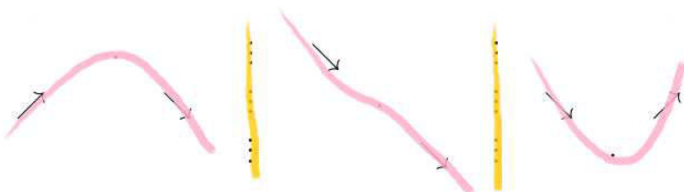
According to the *Intermediate Value Theorem*, at these points and at these points only may the derivative change its sign.

We now list *all* the factors. They are simple enough for us to determine whether and where they change their signs:

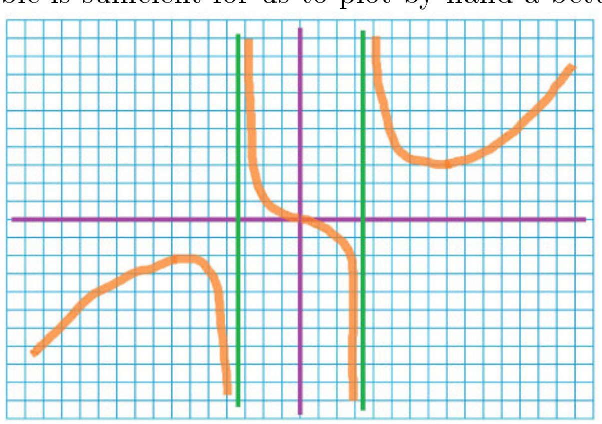
factors	signs											
x^2	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$	$+$	
$x-3$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	$-$	0	$+$	
$x+3$	$-$	0	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	
$(x-\sqrt{3})^2$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	0	$+$	$+$	$+$	
$(x+\sqrt{3})^2$	$+$	$+$	$+$	0	$+$	$+$	$+$	$+$	$+$	$+$	$+$	
domain	$\cdots$	$\bullet$	$\cdots$	$\circ$	$\cdots$	$\bullet$	$\cdots$	$\circ$	$\cdots$	$\bullet$	$\cdots$	$\to x$
$x=$		-3		$-\sqrt{3}$		0		$\sqrt{3}$		3		
f'	$+$	0	$-$	$\vdots$	$-$	0	$-$	$\vdots$	$-$	0	$+$	
		$\cdot$		$\vdots$	$\searrow$			$\vdots$				
f	$\nearrow$		$\searrow$	$\vdots$		$\cdot$		$\vdots$	$\searrow$		$\nearrow$	
				$\vdots$			$\searrow$	$\vdots$		$\cdot$		

Then we go vertically and determine the sign of the derivative using: $+\cdot - = -$, etc. The increasing and decreasing behavior of f is then derived. The extrema are visually classified. We also detect the two vertical asymptotes.

The resulting diagram can serve as a guide for a rough sketch of the graph:



This data collected in the table is sufficient for us to plot by hand a better graph:



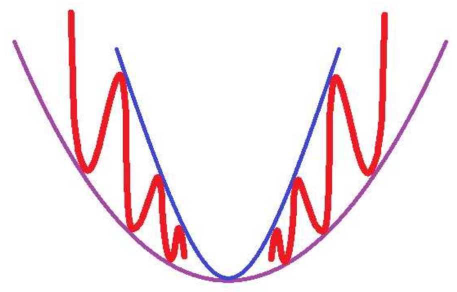
Even then, every part of the graph may lie higher or lower than these. Furthermore, we can't guarantee this concavity without the second derivative!

Example 5.6.4: converse

The converse of the theorem is as follows:

$$\text{A point } c \text{ is max or min of } f \quad \overset{?}{\Longrightarrow} \quad f' \text{ changes its sign at } x=c.$$

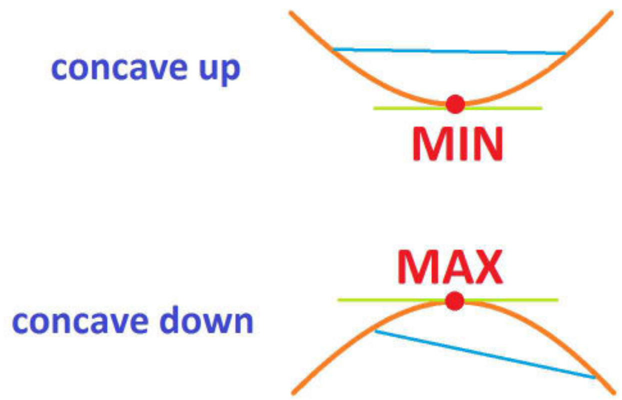
It fails as the following example shows:



Exercise 5.6.5

Devise a formula for the above function and show that the converse of the theorem fails. What do you notice about f'?

We can classify the critical points with the help of just the second derivative at that point:



Theorem 5.6.6: Second Derivative Test

Suppose f is twice differentiable at $x = c$ and suppose $f'(c) = 0$. Then, we have:
- *If $f''(c) \leq 0$, then c is a local max point.*
- *If $f''(c) \geq 0$, then c is a local min point.*

Warning!

If $f''(c) = 0$, then the test fails.

Exercise 5.6.7

Consider these functions:
- $f(x) = x^4$
- $f(x) = x^3$

Example 5.6.8: cubic, continued

Let's consider again:

$$f(x) = x^3 - 3x\,.$$

We could have saved some time with the critical points we found, $x = \pm 1$, if we had forgone the sign analysis of the first derivative f'. Instead, we take the *second* derivative:

$$f'(x) = 3x^2 - 3 \implies f''(x) = 6x\,.$$

Here is the sign analysis of the second derivative:
- $f''(-1) = -6 < 0 \implies x = -1$ is a local max.
- $f''(1) = 6 > 0 \implies x = 1$ is a local min.

Exercise 5.6.9

Carry out a similar analysis for the rational function in the last example.

Example 5.6.10: another rational function

Let
$$f(x) = \frac{x^2}{x^2-1}\,.$$
Here we will simply classify the critical points.

We compute the first derivative:
$$\begin{aligned} f'(x) &= \frac{2x(x^2-1) - x^2 \cdot 2x}{(x^2-1)^2} \\ &= \frac{2x^3 - 2x - 2x^3}{(x^2-1)^2} \\ &= -\frac{2x}{(x^2-1)^2}\,. \end{aligned}$$
We had to simplify and factor this expression in order to facilitate the next differentiation. We can also easily observe the following:
$$f'(x) = 0 \iff x = 0\,.$$
This is the only critical point.

We compute the second derivative:
$$f''(x) = -\frac{2(x^2-1)^2 - 2x \cdot 2(x^2-1)2x}{(x^2-1)^4}\,.$$
No need to simplify! We just need its sign at the critical point:
$$\begin{aligned} f''(0) &= -\frac{2(0^2-1)^2 - 2 \cdot 0 \cdot 2(0^2-1)2 \cdot 0}{(0^2-1)^4} \\ &= -2 < 0\,. \end{aligned}$$
This is a maximum!

Example 5.6.11: another rational function, continued

We continue with
$$f(x) = \frac{x^2}{x^2-1}\,.$$
We now look for a more complete picture.

We start at the bottom:

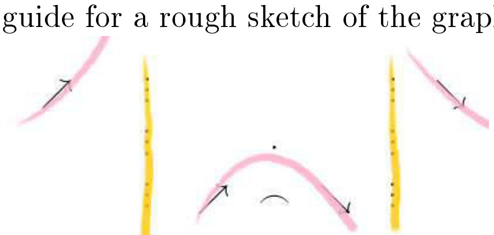
- Domain is all reals except $x = \pm 1$.
- Vertical asymptotes: $x = 1,\ x = -1$.
- Horizontal asymptote: $y = 1$.

Now the derivatives. This time we have to simplify f'' in order to factor:
$$\begin{aligned} f''(x) &= -\frac{2(x^2-1)^2 - 2x \cdot 2(x^2-1)2x}{(x^2-1)^4} \\ &= -\frac{2(x^2-1)(x^2-1-4x^2)}{(x^2-1)^4} \\ &= \frac{2(x-1)(x+1)(3x^2+1)}{(x^2-1)^4}. \end{aligned}$$

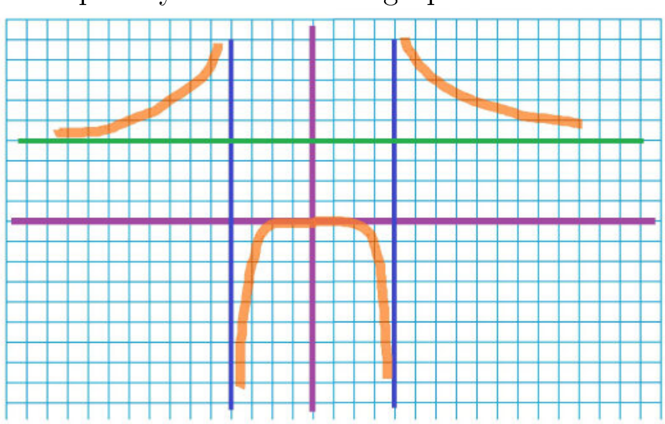
We now need to find the signs of the factors and then the signs of the derivatives. We take note of the domain and only list the factors that may change sign (Volume 1, Chapter 1PC-4):

factors	signs							
$-x$	$+$	$\vdots$	$+$	0	$-$	$\vdots$	$-$	
f'	$+$	$\vdots$	$+$	0	$-$	$\vdots$	$-$	
f	$\nearrow$	$\vdots$	$\nearrow$		$\searrow$	$\vdots$	$\searrow$	According to the Monotonicity Theorem.
$x-1$	$-$	$-$	$-$	$-$	$-$	0	$+$	
$x+1$	$-$	0	$+$	$+$	$+$	$+$	$+$	
f''	$+$	$\vdots$	$-$	$-$	$-$	$\vdots$	$+$	
f	$\smile$	$\vdots$	$\frown$	$\frown$	$\frown$	$\vdots$	$\smile$	According to the Concavity Theorem.
domain	$\cdots$	$\circ$	$\cdots$	$\bullet$	$\cdots$	$\circ$	$\cdots$	$\to x$
$x=$		-1		0		1		
	$\nearrow$	$\vdots$				$\vdots$	$\searrow$	
f		$\vdots$		$\cdot$		$\vdots$		
		$\vdots$	$\nearrow$	$\frown$	$\searrow$	$\vdots$		

The last diagram can serve as a guide for a rough sketch of the graph:



This data is sufficient for us to plot by hand a better graph:



Exercise 5.6.12

Solve a reversed problem: Sketch the graph of a function f that satisfies all of the following:

- $f' > 0$ on $(-\infty, 0)$ and $(1, 2)$;
- $f' < 0$ on $(0, 1)$ and $(2, \infty)$;
- $f'' > 0$ on $(-\infty, -1)$ and $(2, \infty)$;
- $f'' < 0$ on $(-1, 2)$.

5.7. Anti-differentiation: the derivative of what function?

Before the *Mean Value Theorem*, we have only been able to find facts about the derivative from the facts about the function.

This is a short list of familiar facts (Chapter 3):

	info about f		info about f'	
0	f is constant	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	f' is zero.	
1	f is linear	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	f' is constant.	0
2	f is quadratic	$\Longrightarrow$ $\overset{?}{\Longleftarrow}$	f' is linear.	1

The arrows represent differentiation, the reverse arrows represent what we will call *anti-differentiation*.

But are these arrows reversible?

If the derivative of the function is zero, does it mean that the function is constant? Yes, we proved it earlier in this chapter. This time, we have a tool to prove these facts, the *Mean Value Theorem*: If f is continuous on $[a, b]$ and differentiable on (a, b), then

$$\frac{f(b) - f(a)}{b - a} = f'(c)\,,$$

for some c in (a, b).

The theorem will help us with the facts about the function derived from the facts about its derivative.

We start over.

Consider this *obvious* statement about motion:

- "If my speed is zero, I am standing still (and vice versa)."

If a function $y = f(x)$ represents the position, we can restate this mathematically.

Proving the mathematical version of this statement will confirm that our theory matches the reality and the common sense. We follow the same three steps starting with the differences:

Theorem 5.7.1: Constancy vs. Zero Difference

A function defined at the nodes of a partition of interval $[a, b]$ has a zero difference for all secondary nodes in the partition if and only if this function is constant over the nodes of $[a, b]$.

In other words, we have for all secondary nodes:

$$\Delta f = 0 \iff f = \text{ constant}.$$

Proof.

$$\Delta f(c_i) = 0 \implies f(x_i) - f(x_{i-1}) = 0 \implies f(x_i) = f(x_{i-1})\,.$$

Theorem 5.7.2: Constancy vs. Zero Difference Quotient

A function defined at the nodes of a partition of interval $[a,b]$ has a zero difference quotient for all secondary nodes in the partition if and only if this function is constant over the nodes of $[a,b]$.

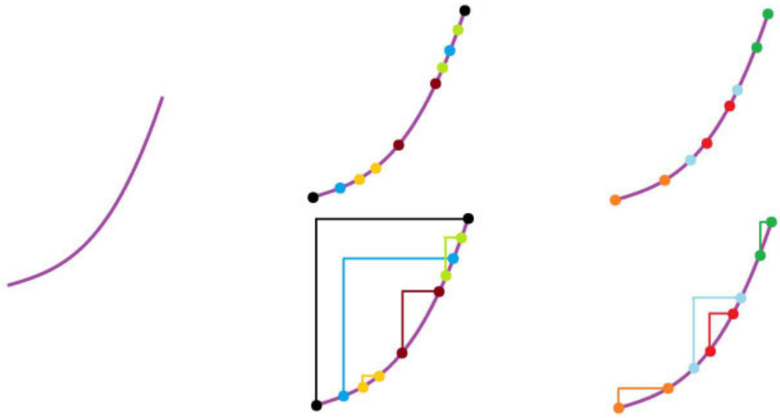
In other words, we have for all secondary nodes:

$$\frac{\Delta f}{\Delta x} = 0 \iff f = \text{ constant}\,.$$

Proof.

$$\frac{\Delta f}{\Delta x}(c_i) = 0 \implies \Delta f(c_i) = 0.$$

The proofs have been so easy because we only had to consider *consecutive* points of the partition. There is no such thing when the function is defined on the whole interval:



Theorem 5.7.3: Constancy vs. Zero Derivative

A differentiable on open interval I function has a zero derivative for all x in I if and only if this function is constant on I.

In other words, we have on I:

$$f' = 0 \iff f = \text{ constant.}$$

Proof.

To prove that f is constant, it suffices to show that

$$f(a) = f(b)\,,$$

for all a, b in I. Assume $a < b$ and use the *Mean Value Theorem* with interval (a,b) inside the interval I:

$$\frac{f(b) - f(a)}{b - a} = f'(c)\,,$$

for some c in (a,b). This is 0 by assumption. Therefore, we have:

$$\frac{f(b) - f(a)}{b - a} = 0\,,$$

for all pairs a, b. Hence

$$f(b) - f(a) = 0\,,$$

or

$$f(a) = f(b)\,.$$

The converse was proven in Chapter 4.

Note that the proof of the Monotonicity Theorems is identical to the one above with each "=" replaced with ">".

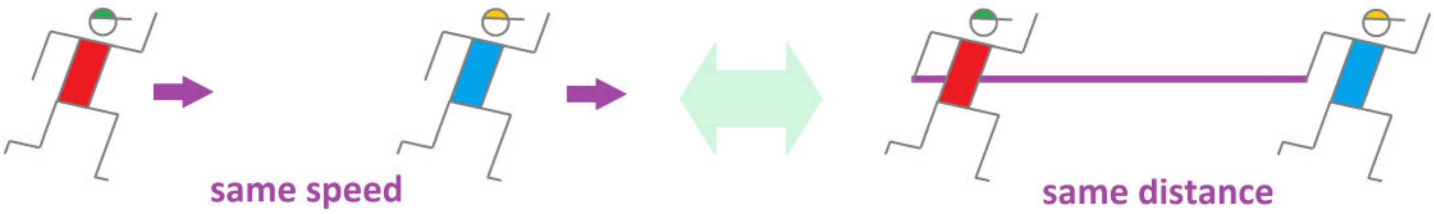
Exercise 5.7.4

What if $f' = 0$ on the union of two open intervals?

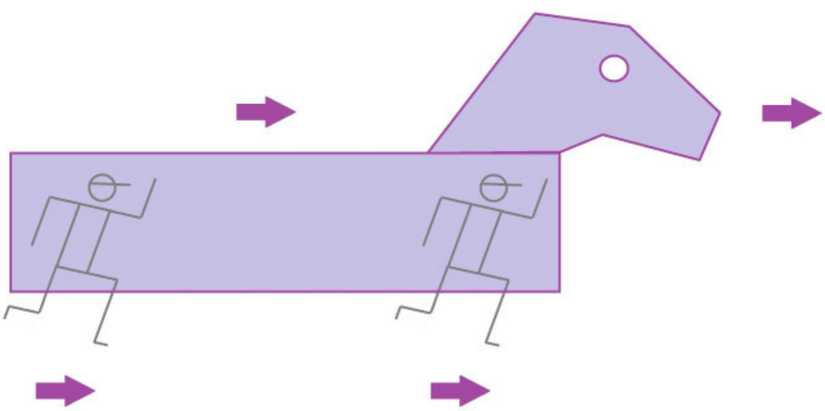
Suppose now that there are *two* runners running with the same speed; what can we say about their mutual locations? They are not, of course, standing still, but they *are* still relative to each other! We have a slightly less obvious fact about motion:

- ▶ "If two runners run with the same speed, the distance between them isn't changing (and vice versa)".

It's as if they are holding the two ends of a pole without pulling or pushing.



The fact remains valid even if they speed up and slow down all the time. They move as if a single body:



Once again, for functions $y = F(x)$ and $y = G(x)$ representing their positions, we can restate this idea mathematically in order to confirm that our theory makes sense. We have to follow the same three steps starting with the differences:

Theorem 5.7.5: Anti-differentiation for Differences

Two functions defined at the nodes of a partition of interval $[a,b]$ have the same differences if and only if theydiffer by a constant.

In other words, we have:

$$\Delta F(c) = \Delta G(c) \iff F(x) - G(x) = \text{ constant.}$$

Theorem 5.7.6: Anti-differentiation for Different Quotients

Two functions defined at the nodes of a partition of interval $[a,b]$ have the same difference quotient if and only if they differ by a constant.

In other words, we have:

$$\frac{\Delta F}{\Delta x}(c) = \frac{\Delta G}{\Delta x}(c) \iff F(x) - G(x) = \text{ constant.}$$

Once again, this has been easy because we only had to consider consecutive points of the partition. For a function defined on the whole interval, we'd need the Mean Value Theorem. Here instead, the theorem about the constancy of a function with a zero derivative is used:

Theorem 5.7.7: Anti-differentiation for Derivatives

Two differentiable on open interval I functions have the same derivative if and only if they differ by a constant.

In other words, we have:

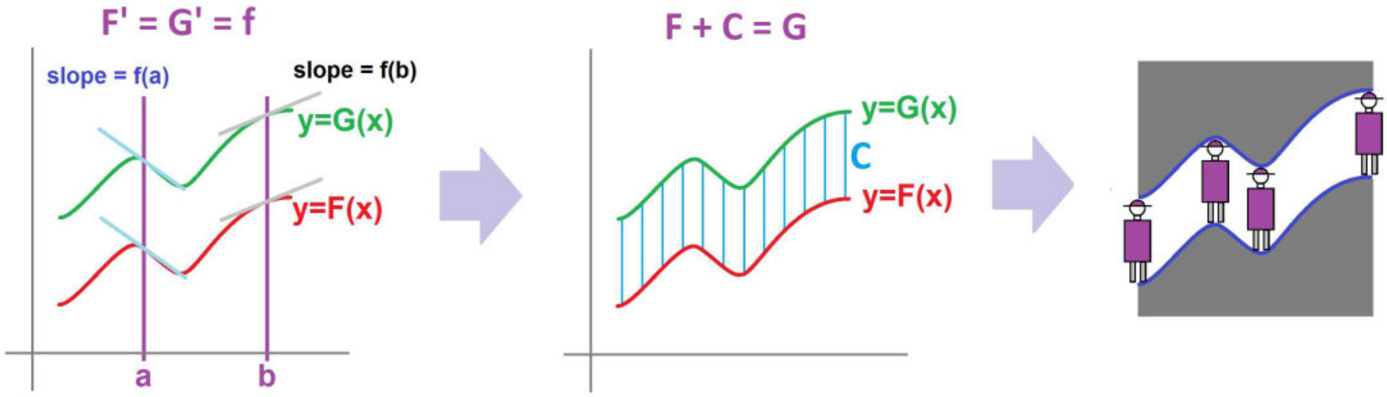
$$F'(x) = G'(x) \iff F(x) - G(x) = \text{ constant.}$$

Proof.

Define
$$h(x) = F(x) - G(x)\,.$$
Then, by the Sum Rule, we have:
$$h'(x) = (F(x) - G(x))' = F'(x) - G'(x) = 0\,,$$
for all x. Then h is constant, by the *Constancy vs. Zero Derivative Theorem.* The converse was proven in Chapter 4.

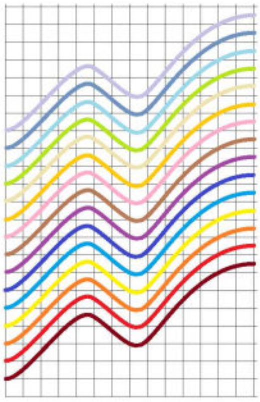
In addition to the motion interpretation, there is also one in terms of geometry. The theorem says:

▶ If the graphs of $y = F(x)$ and $y = G(x)$ have parallel tangent lines for every value of x, then the graph of F is a vertical shift of the graph of G (and vice versa).



We can understand this idea if we imagine a tunnel and a person whose head is touching the ceiling. If the ceiling is sloped down, should he be concerned about hitting his head? Not if the floor is sloped down as much! In other words, if the slope of the tunnel's top is equal to the slope of the bottom at every location, then the height of the tunnel remains the same throughout its length.

There are infinitely many functions with the same derivative $F' = f$:



So, even if we can recover the function F from its derivative F', there are many others with the same derivative, such as $G = F + C$ for any constant real number C. Are there others? Not according to the theorem.

Warning!

It's only true when the domain is an interval.

Based on the theorem, we can now update this list of simple but important facts:

	info about f		info about f'	
0	f is constant.	$\Longleftrightarrow$	f' is zero.	
1	f is linear.	$\Longleftrightarrow$	f' is constant.	0
2	f is quadratic.	$\Longleftrightarrow$	f' is linear.	1
...				...

We notice a pattern: f is a polynomial of degree n if and only if f' is a polynomial of degree $n-1$.

We use the last two facts to justify our analysis of *free fall*:

Functions of time	
The acceleration is constant.	$\Longrightarrow$
The velocity is linear.	$\Longrightarrow$
The location is quadratic.	

Exercise 5.7.8

Derive that the trajectory of a ball thrown under an angle is a parabola.

5.8. *Antiderivatives*

Every good problem has a "reversed" counterpart, usually harder. This one is for differentiation:

Definition 5.8.1: antiderivative on partition

Suppose a function f is defined on the secondary nodes of a partition of a closed interval I. Then a function F defined on the nodes of the partition that satisfies

the equation:

$$\boxed{\frac{\Delta F}{\Delta x}(c) = f(c)}$$

for all secondary nodes c, is called an *antiderivative* of f.

In the previous chapters, we found a recursive solution of this equation by solving the equation for the difference quotient as in the case of position from velocity:

$$v_n = \frac{p_{n+1} - p_n}{\Delta t} \implies p_{n+1} = p_n + v_n \Delta t\,.$$

It's pure algebra!

The continuous case is by far more complex:

Definition 5.8.2: antiderivative on interval

Suppose a function f is defined on an open interval I. Then a differentiable function F defined on I that satisfies the equation:

$$\boxed{\frac{dF}{dx}(x) = f(x)}$$

for all x, is called an *antiderivative* of f.

We use "an" because there are many antiderivatives for each function. As we know from the *Anti-differentiation Theorem*, if F is an antiderivative of f, then so is $F + C$, where C is any constant, while the converse is true only for functions defined on *intervals*.

We can think of the definition as an *equation*, an equation for functions:

Given $f,$ solve for F	Given $f,$ solve for F
$\dfrac{\Delta F}{\Delta x} = f$	$\dfrac{dF}{dx} = f$

Example 5.8.3: x^2

For example, the task may be to find a function F that satisfies the equation:

$$F'(x) = x^2\,.$$

As it is often the case with equations, there is no formula that directly gives a solution. There is no limit either. The initial idea is then to try to recall *prior experiences with differentiation*. Even if you've never finished differentiating a function with x^2 as the answer, you might have done something close:

$$(x^3)' = 3x^2\,.$$

Almost! To get rid of 3, *try* to divide: maybe $x^3/3$? Let's test:

$$(x^3/3)' = (x^3)/3 = 3x^2/3 = x^2\,.$$

As it is often the case with equations, trial-and-error might just work!

We can restate the *Anti-differentiation Theorem* as follows:

Corollary 5.8.4: Set of Antiderivatives

Suppose F is any antiderivative of a function f defined on an open interval. Then the set of all of its antiderivatives is

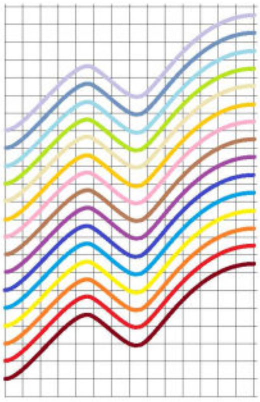
$$\{F + C : C \text{ real}\}.$$

As it is often the case with equations, there seems to be many (infinitely many) solutions. But a very important conclusion is that *it suffices to find just one anti-derivative!*

Warning!

The formula $F + C$ works only when the function is defined on an interval.

This is what this set of functions would look like:



It is the solution set of the equation and it may be called *the* antiderivative.

Exercise 5.8.5: x^3

Find all F that satisfy the equation:

$$F'(x) = x^3.$$

Exercise 5.8.6

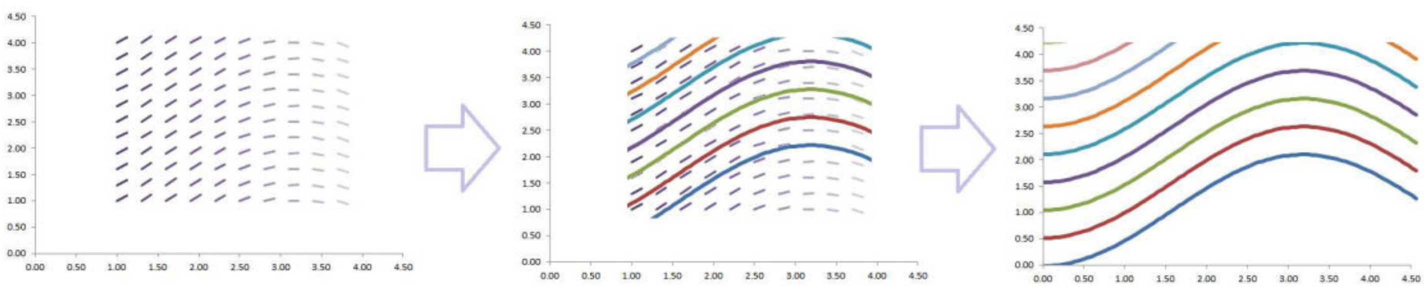
Suppose a function f is defined on an open interval I. Prove that the following:

1. The graphs of two different antiderivatives of f never intersect.
2. For every point (x, y) with x within I, there is an antiderivative of f the graph of which passes through it.

The problem then becomes the one of finding a single function F, either

- from its difference quotient $\dfrac{\Delta F}{\Delta x}$, or
- from its derivative $\dfrac{dF}{dx}$.

In other words, we reconstruct the function from a "field of slopes":

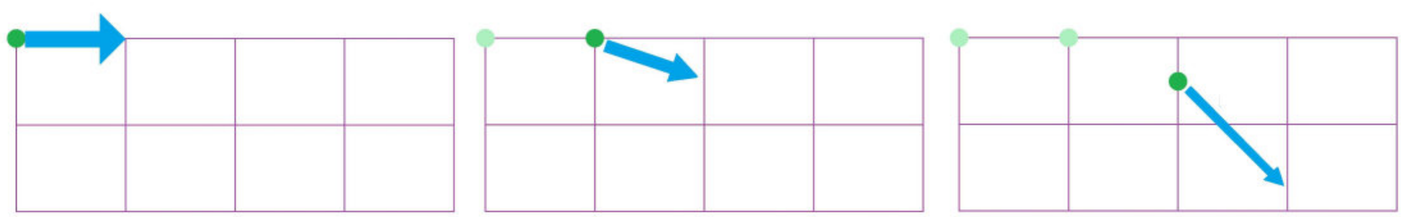


One can imagine a flowing liquid with the direction known at every location. How do we find the path of a particular particle? The process of reconstructing a function, F, from its derivative, f, is called *anti-differentiation*.

The anti-differentiation problem has been solved on several occasions for the former, discrete case – velocity from acceleration and location from velocity – via these recursive formulas:

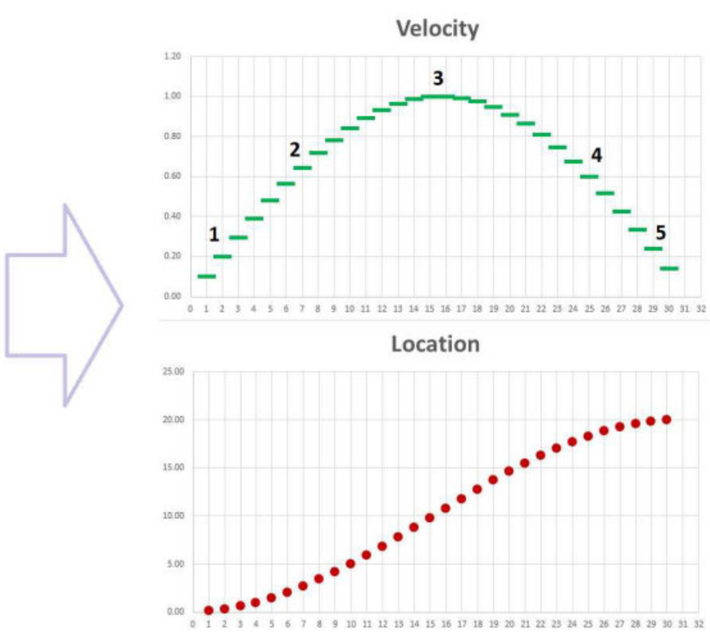
$$F(x_{n+1}) = F(x_n) + f(c_n)\Delta x_n\,.$$

For each location, we look up the velocity, find the next location, and repeat:



If the nodes of the partition are close enough to each other, these points form curves:

time min	velocity miles/min	location miles
0		0.00
1	0.10	0.10
2	0.20	0.30
3	0.30	0.59
4	0.39	0.98
5	0.48	1.46
6	0.56	2.03
7	0.64	2.67
8	0.72	3.39
9	0.78	4.17
10	0.84	5.01
11	0.89	5.91
12	0.93	6.84
13	0.96	7.80
14	0.99	8.79
15	1.00	9.78
16	1.00	10.78
17	0.99	11.77
18	0.97	12.75
19	0.95	13.70
20	0.91	14.60
21	0.86	15.47
22	0.81	16.28
23	0.75	17.02
24	0.68	17.70
25	0.60	18.30
26	0.52	18.81
27	0.43	19.24
28	0.33	19.57
29	0.24	19.81
30	0.14	19.95



For the latter, continuous case, this is a challenging problem: How does one plot a curve that follows these – infinitely many – tangents?

To begin with, we just try to *reverse differentiation*. We will try to construct a theory of anti-differentiation that matches – to the degree possible – that of differentiation.

Here is a short *list of derivatives* of functions (for all x for which the function is differentiable):

function $\longrightarrow$	derivative
x^r	rx^{r-1}
$\ln x$	$\frac{1}{x}$
e^x	e^x
$\sin x$	$\cos x$
$\cos x$	$-\sin x$
antiderivative $\longleftarrow$	function

To find antiderivatives, reverse the order:

- ▶ Read each line from right to left!

Example 5.8.7: $\sin$ and $\cos$

What is an antiderivative of $\cos x$? We need to solve for F:

$$F'(x) = \cos x\,.$$

Just find $\cos x$ in the right column. The corresponding function on the left is $\sin x$. That's the answer: $F(x) = \sin x$!

What is an antiderivative of $\sin x$? Solve:

$$F'(x) = \sin x\,.$$

Just find $\sin x$ in the right column. It's not there... but $-\sin x$ is! The corresponding function on the left is $\cos x$. Then (according to the Constant Multiple Rule) the solution must be $-\sin x$.

It is that simple! We just may need some tweaking to make the formulas about to emerge as easy to apply as the original ones.

For example, let's find antiderivatives of x^n. Use the *Power Formula* for differentiation (the first row), divide by r, and apply the *Constant Multiple Rule*:

$$(x^r)' = rx^{r-1} \implies \frac{1}{r}(x^r)' = x^{r-1} \implies \left(\frac{1}{r}x^r\right)' = x^{r-1}\,.$$

We then simplify the right-hand side by setting $r-1=s$:

$$\left(\frac{1}{s+1}x^{s+1}\right)' = x^s\,.$$

Example 5.8.8: $\sqrt[3]{\ }$

Solve for F:

$$F'(x) = \sqrt[3]{x} = x^{1/3}\,.$$

We choose $s = 1/3$ in the formula. Then, we have:

$$F(x) = \frac{1}{1/3+1}x^{1/3+1} = \frac{3}{4}x^{4/3} + C\,.$$

We make the right-hand side the left-hand side and we have the *Power Formula* for anti-differentiation. We have the following:

$$\text{An antiderivative of } x^s \text{ is } \frac{1}{s+1}x^{s+1}, \text{ provided } s \neq -1.$$

But what if $s = -1$? Then we read the answer from the next line in the table:

$$\text{An antiderivative of } x^{-1} \text{ is } \ln|x|.$$

So, that the rule "the derivative of a power function is a power function of degree 1 lower" has an exception, the 0-power, and the rule "the antiderivative of a power function is a power function of degree 1 higher" has an exception too.

Taking the rest of these rows, we have a *list of antiderivatives* of functions, on open intervals:

function $\longrightarrow$	antiderivative
x^s	$\frac{1}{s+1}x^{s+1}, \quad s \neq -1$
$\frac{1}{x}$	$\ln|x|$
e^x	e^x
$\sin x$	$-\cos x$
$\cos x$	$\sin x$
derivative $\longleftarrow$	function

Example 5.8.9: domains

Each formula is only valid on an open interval on which the antiderivative is defined. For example, we interpret the second row as follows:

- $\ln(x)$ is an antiderivative of $\frac{1}{x}$ on the interval $(0, +\infty)$, and
- $\ln(-x)$ is an antiderivative of $\frac{1}{x}$ on the interval $(-\infty, 0)$.

Next, we will need the *rules of anti-differentiation*.

First, consider the *Sum Rule for Derivatives* (Chapter 4): The derivative of the sum is the sum of the derivatives; i.e.,

$$(f+g)' = f' + g'\,.$$

Let's read that formula from right to left:

Theorem 5.8.10: Sum Rule for Antiderivatives

An antiderivative of the sum is the sum of the antiderivatives.

In other words, we have:

if

F *is an antiderivative of* f *and*

G *is an antiderivative of* g,

then

$F + G$ *is an antiderivative of* $f + g$.

Proof.

We apply the *Sum Rule For Derivatives* to confirm:

$$(F(x)+G(x))' = F'(x)+G'(x) = f(x)+g(x)\,.$$

Exercise 5.8.11

What about the converse?

Example 5.8.12: sums

Solve for F:

$$F'(x) = x^2 + \sin x\,.$$

The equation is solved by solving the following two equations:

$$\text{for } \ G:\ G'(x) = x^2 \ \text{ and for } \ H:\ H'(x) = \sin x\,.$$

The solutions are found in the table:

$$G(x) = \frac{1}{3}x^3 \ \text{ and } \ H(x) = -\cos x\,.$$

According to the theorem, we have an answer:

$$F(x) = \frac{1}{3}x^3 - \cos x\,.$$

Exercise 5.8.13

Using this rule, find an antiderivative of $\ln x^2$.

Compare:

The	derivative	of the sum is the sum of the	derivatives.
The	antiderivative	of the sum is the sum of the	antiderivatives.

Warning!

The mismatch between "*an* antiderivative" and "*the* antiderivative" isn't accidental.

Similarly, consider the *Constant Multiple Rule for Derivatives* (Chapter 4): The derivative of a multiple is the multiple of the derivative; i.e.,

$$(cf)' = cf'\,.$$

Let's read that formula from right to left:

Theorem 5.8.14: Constant Multiple Rule for Antiderivatives

An antiderivative of a multiple is the multiple of the antiderivative.

In other words, we have:
if
F *is and antiderivative of* f *and*
c *is a constant,*
then
cF *is an antiderivative of* cf.

Proof.

We apply the *Constant Multiple Rule For Derivatives*:

$$(cF(x))' = cF'(x) = cf(x)\,.$$

Exercise 5.8.15

What about the converse?

Example 5.8.16: constant multiples

Solve for F:

$$F'(x) = 3\sin x\,.$$

We solve the equation by solving the following equation:

$$\text{for}\quad G:\ G'(x) = \sin x\,.$$

The solution is found in the table:

$$G(x) = -\cos x\,.$$

According to the theorem, we have an answer:

$$F(x) = 3(-\cos x)\,.$$

Exercise 5.8.17

Using this rule, find an antiderivative of e^{x+3}.

Compare:

The	derivative	of a constant multiple is the constant multiple of the	derivative.
The	antiderivative	of a constant multiple is the constant multiple of the	antiderivative.

We can handle compositions just as easily but only when one of the functions is linear. We read the *Linear Chain Rule* (Chapter 4) from right to left:

$$\frac{d}{dx}f(mx+b) = mf'(mx+b)\,,$$

producing the following:

Theorem 5.8.18: Linear Composition Rule for Antiderivatives

If

F *is an antiderivative of* f *and*

$m \neq 0,\ b$ *are constants,*

then

$\dfrac{1}{m}F(mx+b)$ *is an antiderivative of* $f(mx+b)$.

Proof.

We apply the *Linear Chain Rule* and the *Constant Multiple Rule for Differentiation*:

$$\left(\frac{1}{a}F(mx+b)\right)' = \frac{1}{m}\left(F(mx+b)\right)' = \frac{1}{m}mF'(mx+b) = F'(mx+b) = f(mx+b)\,.$$

Exercise 5.8.19

Using this rule, find an antiderivative of e^{2x+1}.

Exercise 5.8.20

What about the other Linear Chain Rule (Chapter 4)?

As we know from the *Anti-differentiation Theorem*, every antiderivative of a function comes with infinitely many others:

$$F \ \to \ F + C \ \text{ for every real } C\,,$$

on every open interval. Together they form *the* antiderivative of the function.

Example 5.8.21: free fall

We can make our analysis of *free fall* more specific:

Functions of time		
The acceleration is constant:	$a = -g.$	$\Longrightarrow$
The velocity is linear:	$v = -gt + C.$	$\Longrightarrow$
The location is quadratic:	$p = -gt^2/2 + Ct + K.$	

The constants C and K cover all possible trips of the ball.

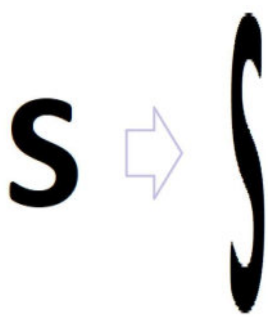
Below is the standard notation for *the* antiderivative of a function f:

Antiderivative

$$\int f\, dx$$

Here, $\displaystyle\int$ is referred to as the "integral sign".

It looks like a stretched letter "S", which stands for "summation":



The summation being referred to here is the one in the recursive formula for discrete anti-derivative:

$$F(x_{n+1}) = F(x_n) \, + \, f(c_n)\Delta x_n\,.$$

The notation idea for anti-differentiation matches that for differentiation when written in a certain way:

$$\frac{d}{dx}\Big(\ 3x^2 + \cos x\ \Big) \ = 6x - \sin x\,.$$

We see the "parentheses" of this function of functions. This is how the new notation is deconstructed:

Antiderivative

$$\underset{\substack{\uparrow\\ \text{left and}}}{\int (} \; \overset{\substack{\text{a specific function}\\ \downarrow}}{3x^2} + \cos x \underset{\substack{\uparrow\\ \text{right "parentheses"}}}{) dx} = \overset{\substack{\text{a specific function}\\ \downarrow}}{x^3 + \sin x} + \underset{\substack{\uparrow\\ \text{indicates that there are others}}}{C}$$

This is how we rewrite the above list:

$$\begin{aligned}
\int x^s\, dx &= \frac{1}{s+1}x^{s+1} + C, \quad \text{for } s \neq -1\\
\int \frac{1}{x}\, dx &= \ln x + C\\
\int e^x\, dx &= e^x + C\\
\int \sin x\, dx &= -\cos x + C\\
\int \cos x\, dx &= \sin x + C
\end{aligned}$$

We restate the rules too.

Sum Rule:

$$\int (f+g)\, dx = \int f\, dx + \int g\, dx$$

Constant Multiple Rule:

$$\int (cf)\, dx = c\int f\, dx$$

Linear Composition Rule:

$$\int f(mx+b)\, dx = \frac{1}{m}\int f(t)\, dt\Big|_{t=mx+b}$$

With these rules, when applicable, anti-differentiation is very similar to differentiation.

Example 5.8.22: rules of anti-differentiation

Find an antiderivative of

$$3x^2 + 5e^x + \cos x\,.$$

One can imagine what he'd do to differentiate and then follow the same steps but with the anti-differentiation formulas and rules used instead.

Differentiation:

$$\begin{aligned}
(3x^2 + 5e^x + \cos x)' &= (3x^2)' + 5(e^x)' + (3\sin x)' && \text{SR}\\
&= 3(x^2)' + 5(e^x)' + 3(\sin x)' && \text{CMR}\\
&= 3\cdot 6x + 5e^x + 3\cos x\,. && \text{Table}
\end{aligned}$$

Anti-differentiation:

$$\begin{aligned}\int (3x^2+5e^x+\cos x)'dx &= \int (3x^2)\,dx + \int 5(e^x)\,dx + \int (3\sin x)dx && \text{SR}\\ &= 3\int (x^2)\,dx + 5\int (e^x)\,dx + 3\int (\sin x)dx && \text{CMR}\\ &= 3\cdot x^3/3 + 5e^x + 3(-\cos x) + C\,. && \text{Table}\end{aligned}$$

Just as when solving equations, we can easily confirm that our computations were correct, by substitution. In this case, we *differentiate the antiderivative*:

$$\begin{aligned}(x^3+5e^x+3\sin x)' &= (x^3)' + 5(e^x)' + (3\sin x)'\\ &= 3x^2 + 5e^x + 3\cos x\,.\end{aligned}$$

This is the original function! The answer checks out.

Exercise 5.8.23

Find the antiderivative of $5e^{3x+2} - e^e$.

Below, we have these two diagrams to illustrate the interaction of antiderivatives with algebra:

$$\begin{tikzcd} f,g \arrow[d, "+"] & f',g' \arrow[l, "\int"'] \arrow[d, "+"] \\ f+g & f'+g' \arrow[l, "\int"'] \end{tikzcd} \qquad \begin{tikzcd} f \arrow[d, "\cdot c"] & f' \arrow[l, "\int"'] \arrow[d, "\cdot c"] \\ cf & cf' \arrow[l, "\int"'] \end{tikzcd}$$

The arrows of differentiation are reversed! We start with a pair of functions at top right, then we proceed in two ways:

- left: anti-differentiate them, then down: add the results; or
- down: add them, then left: anti-differentiate the results.

The result is the same!

So far, this is very similar to differentiation. The strategy is the same: divide and concur. Split addition with the Sum Rule, then factor out the coefficients with the Constant Multiple Rule, then apply the table results to these pieces.

Unfortunately, this is where the similarities stop.

There is no analog of the Product Rule for the Derivatives:

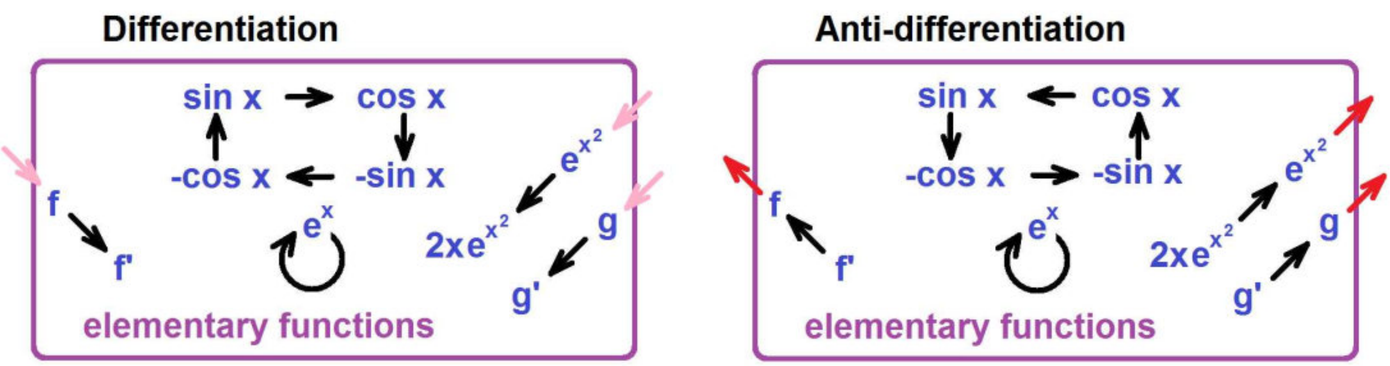
The	derivative	of the product of two functions	*can*	
		be expressed in terms of their	derivatives	and the functions themselves.
The	antiderivative	of the product of two functions	*cannot*	
		be expressed in terms of their	antiderivatives	and the functions themselves.

Similarly there is no Quotient Rule, nor the Chain Rule, for anti-differentiation.

This difference has profound consequences. We can start with just these functions:

$$x^s,\ \sin x, e^x\,.$$

Then – by applying the four algebraic operations, composition, and inverting – we can construct a great variety of functions. Let's call them "elementary functions". Because of the way they are constructed, *all* of them can be easily differentiated with the rules of differentiation thus producing other elementary functions (left):



However, contrary to what the above list might suggest, anti-differentiation will often take us outside of the realm of elementary functions (right). For example, a new function, called the *Gauss error function*, must be created for this important antiderivative:

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int e^{-x^2}\,dx\,.$$

The result will be the same if we *exclude* from the "elementary functions" either the trigonometric functions or the exponent. The result will be the same if we *include* more functions to the list.

With the help of the *Anti-differentiation Theorem*, we can claim that we have found *all* antiderivatives of these functions on our list, over open intervals within the domains of antiderivatives, i.e., *the* antiderivative of the corresponding function:

Definition 5.8.24: general antiderivative

For a given function f, *the general antiderivative of f over open interval I* is defined by:

$$\int f\,dx=F(x)+C\,,$$

where F is any antiderivative of f on I, i.e., $F'=f$, understood as a collection of all such functions over all possible real numbers C. This collection is also called *the indefinite integral* of f.

These diagrams illustrate how differentiation and anti-differentiation undo each other:

$$\begin{array}{ccccccccc}
f & \to & \boxed{\int \square\,dx} & \to & F & \to & \boxed{\frac{d}{dx}\square} & \to & f\\
F & \to & \boxed{\frac{d}{dx}\square} & \to & f & \to & \boxed{\int \square\,dx} & \to & F+C
\end{array}$$

In the second row, we see that $\frac{d}{dx}$ isn't invertible (Volume 1, Chapter 1PC-3).

Example 5.8.25: how many antiderivatives?

There are infinitely many antiderivatives but there is more to it. Let's take a more careful look at one line on the list:

$$\int\frac{1}{x}\,dx \overset{???}{=\!=} \ln|x|+C,\quad x\neq 0\,.$$

This formula is intended to mean the following:

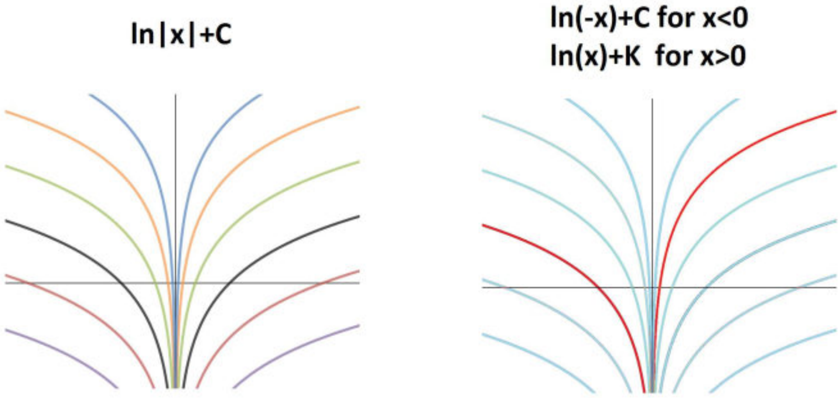
1. We have captured infinitely many – one for each real number C – antiderivatives.
2. We have captured *all* of them.

However, the *Anti-differentiation Theorem* applies only to *one interval at a time*. Meanwhile, the

domain of $1/x$ consists of two rays $(-\infty, 0)$ and $(0, +\infty)$. As a result, we solve this problem *separately* on either of the two intervals. Then the antiderivatives of $1/x$ are:

- $\ln(-x) + C$ on $(-\infty, 0)$, and
- $\ln(x) + C$ on $(0, +\infty)$.

But if now we were to combine each of these pairs of functions into one, F, defined on $(-\infty, 0) \cup (0, +\infty)$, we would realize that, every time, the two constants might be different: After all, they have nothing to do with each other! We illustrate the wrong (incomplete) answer on the left, and the right one on the right:



The image on the left, as well as the formula we started with, might suggest that all of the function's antiderivatives are even functions. The image on the right shows a single antiderivative (in red) but its two branches don't have to match! Algebraically, the antiderivative of $\frac{1}{x}$ – on the whole domain – is given by this piecewise-defined function:

$$F(x) = \begin{cases} \ln(-x) + C & \text{for } x \text{ in } (-\infty, 0), \\ \ln(x) + K & \text{for } x \text{ in } (0, +\infty). \end{cases}$$

It has *two* parameters instead of the usual one.

Exercise 5.8.26

Verify that this is an antiderivative of $1/x$.

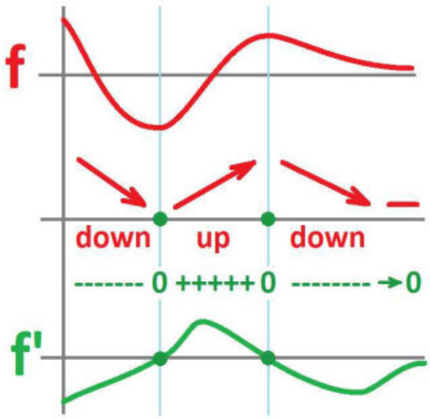
Exercise 5.8.27

In a similar fashion, examine the Power Formula above for $s < -1$.

Example 5.8.28: find graphs

The antiderivatives on our list were discovered by reading the results of differentiation backwards. We can do the same for graphs.

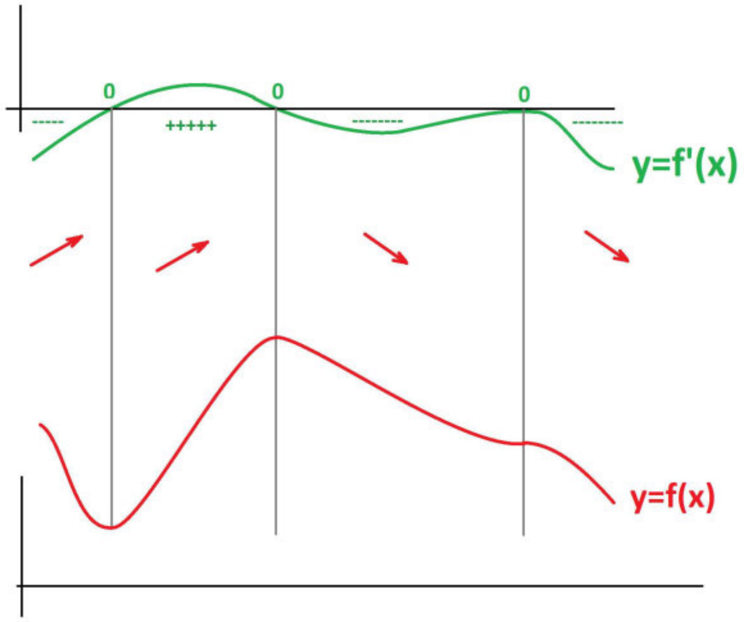
Below, the derivative's graph (green) was found from the graph of the function (red) by looking at the monotonic behavior of f: either $f' > 0$ or $f' < 0$, and local extreme points: $f' = 0$.



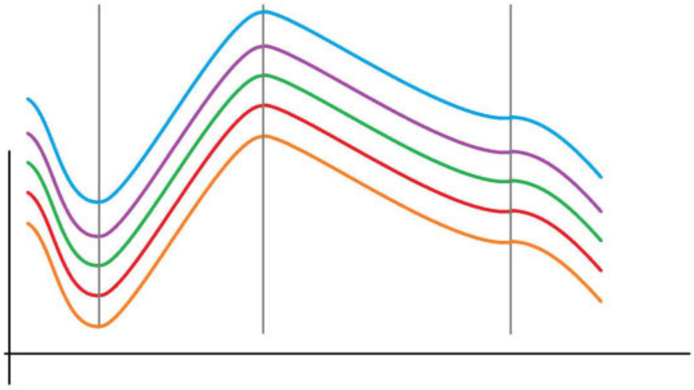
In summary, we look at $\searrow$, $\nearrow$ of f to find $+$, $-$ of f'.

In reverse, we look at $+$, $-$ of f' to find $\searrow$, $\nearrow$ of f.

Here is an example how the graph of f is found from the graph of its derivative:



To show the *answer*, we have to show a multitude of antiderivatives:



Exercise 5.8.29

Find the inflection points.

This study continues in Volume 3, Chapter 3IC-1.

We have been following (and continue) to follow one of the most crucial ideas of the book:

$$\boxed{\lim_{\Delta x\to 0}\left(\begin{array}{c}\text{discrete}\\\text{calculus}\end{array}\right) = \text{calculus}}$$

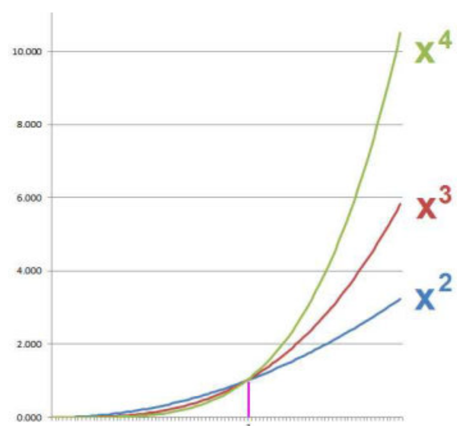
Chapter 6: What we can do with calculus

Contents

6.1. Magnitudes of functions; L'Hopital's Rule

How does one compare the magnitude or the "strength" of two functions on a large scale?

For example, let's take a look at the *power functions*, say x vs. x^2, or x^2 vs. x^3:



The graph of the latter overtakes that of the former and then becomes incomparable higher. Then the answer is obvious:

- ▶ The higher the power, the stronger the function.

A similar rule applies to *polynomials*:

- ▶ The higher the degree, the stronger the polynomial.

But what does "stronger" mean?

To compare two functions, the first thing to look at is the *difference* of the two functions. However, by this standard,

$$x^2 \quad \text{and} \quad x^2 + x$$

are infinitely far apart! At least, that's what the limit of the difference is at infinity.

Instead, we look at is the *ratio* of the two functions. Consider again:

$$x^2 \quad \text{and} \quad x^2 + x\,.$$

By this standard, they are the same! At least, that's what the limit of this ratio is at infinity:

$$\frac{x^2}{x^2 + x} \to 1 \quad \text{as } x \to \infty\,.$$

Example 6.1.1: horizontal asymptotes

We encountered this problem indirectly when computing limits at infinity of rational functions (Chapter 2).

Recall the method we used: We divide both numerator and denominator by the highest available power. For example:

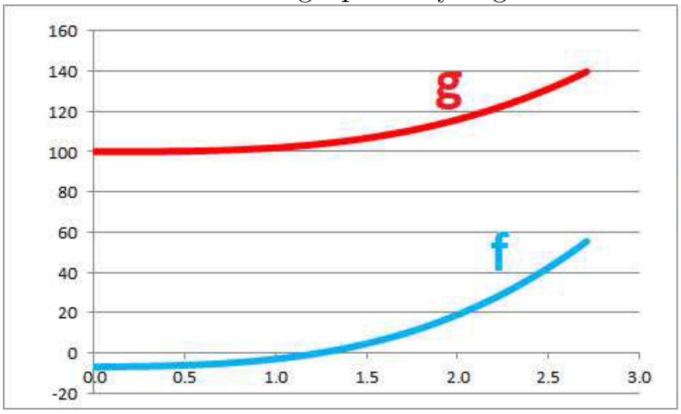
$$\begin{aligned}\frac{3x^3 + x - 7}{2x^3 + 100} &= \frac{3x^3 + x - 7}{2x^3 + 100}\frac{/x^3}{/x^3} \\ &= \frac{3 + x^{-2} - 7x^{-3}}{2 + 100x^{-3}} \\ &\to \frac{3}{2} \qquad\qquad \text{as } x \to \infty\,.\end{aligned}$$

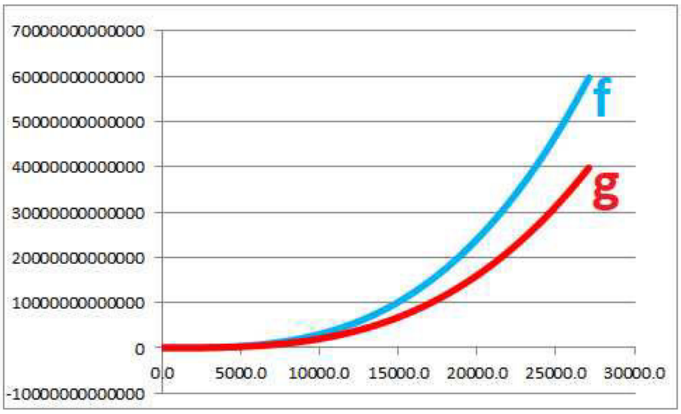
We conclude that these two functions,

$$f(x) = 3x^3 + x - 7 \quad \text{and} \quad g(x) = 2x^3 + 100\,,$$

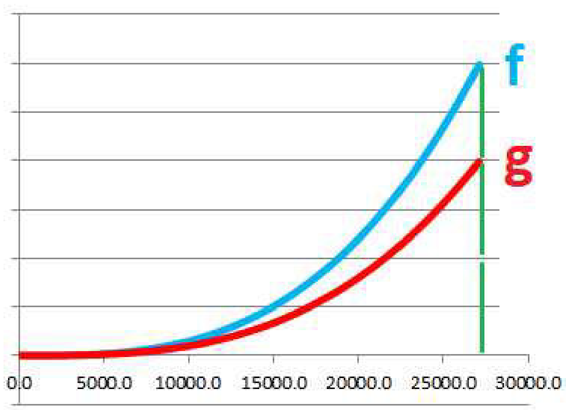
have the "same power" at infinity.

We can see how their graphs stay together below:





Even though the difference goes to infinity, the proportion remains visibly 3 to 2:

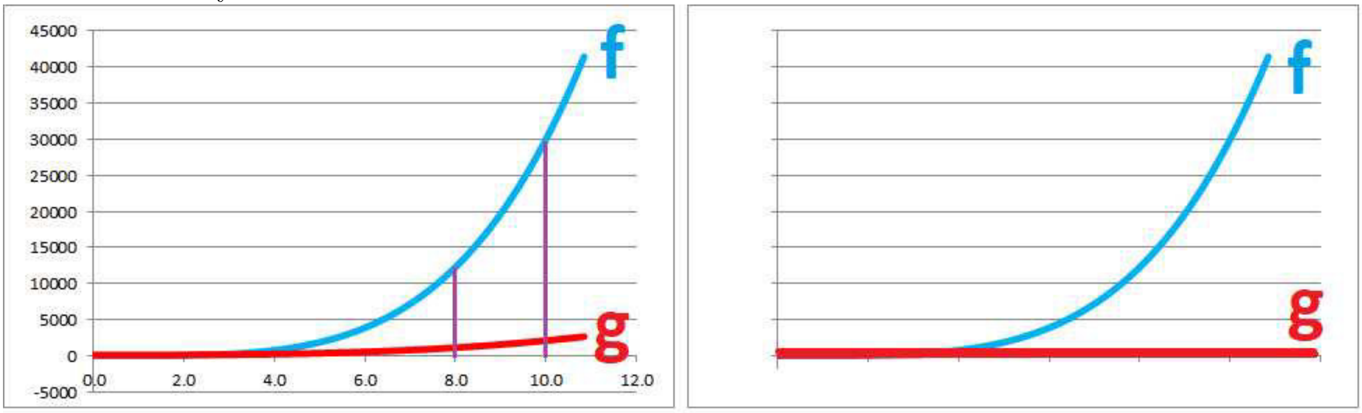


We recognize a vertical stretch! The two functions behave roughly the same.

However, if we replace the power of the numerator with 4, we see this:

$$\begin{aligned}\frac{f(x)}{g(x)} &= \frac{3x^4 + x - 7}{2x^3 + 100} \\ &= \frac{3x^4 + x - 7}{2x^3 + 100}\frac{/x^4}{/x^4} \\ &= \frac{3 + x^{-3} - 7x^{-4}}{2x^{-1} + 100x^{-4}} \\ &\to \infty \qquad\qquad \text{as } x \to \infty\,.\end{aligned}$$

We conclude that $3x^4 + x - 7$ is "stronger" at infinity than $2x^3 + 100$. We can see how quickly the former runs away below:



This isn't a stretch! Further enlarging the domain will cause the plot to shrink horizontally and the smaller function will disappear into the x-axis.

Thus to compare two functions f and g, at infinity or at a point, we form a fraction from them and evaluate the limit of the ratio:

$$\lim_{\square} \frac{f(x)}{g(x)} =?$$

No matter what's in the box, the definition is the same:

Definition 6.1.2: orders of magnitude of functions

If the limit below, with

$$x \to \pm\infty \quad \text{or} \quad x \to a^{\pm}\,,$$

is infinite or its reciprocal is zero,

$$\lim_{\square} \left|\frac{f(x)}{g(x)}\right| = \infty \quad \text{or} \quad \lim_{\square} \left|\frac{g(x)}{f(x)}\right| = 0\,,$$

we say that f is of *higher order than* g. When this limit is a non-zero number, we say that f and g are of *the same order.*

We use the following notation for the former case:

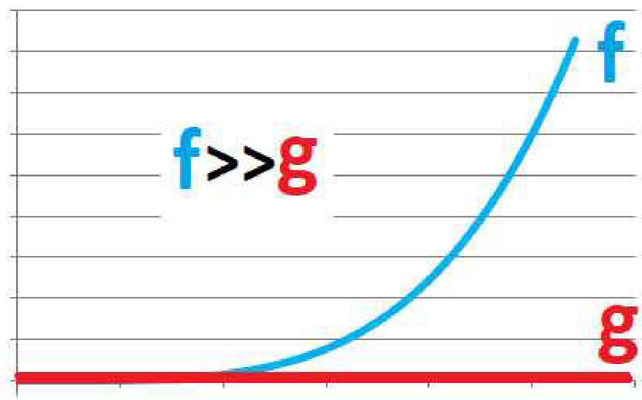
Higher order

$$f >> g$$

$$g = o(f)$$

The latter reads "little o".

When magnified, their graphs look like those in the last example:

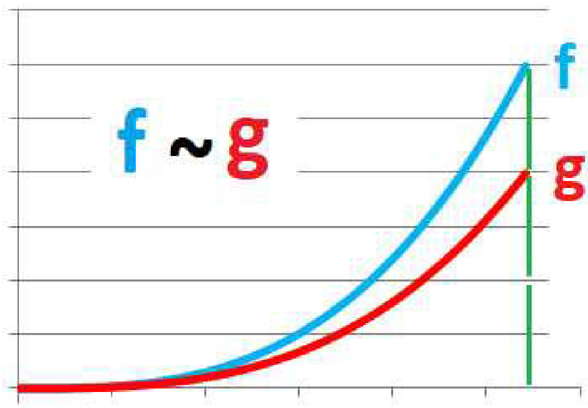


We use the following notation for the latter case:

Equal order

$$f \sim g$$

When magnified, either graph looks like a stretched version of the other:



One of the most important uses of the notation is in the *definition of the derivative*:

$$f'(a) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}\,.$$

Using the Sum Rule, it can be rewritten as follows:

$$\lim_{\Delta x \to 0} \left[\frac{\Delta y}{\Delta x} - f'(a) \right] = 0\,,$$

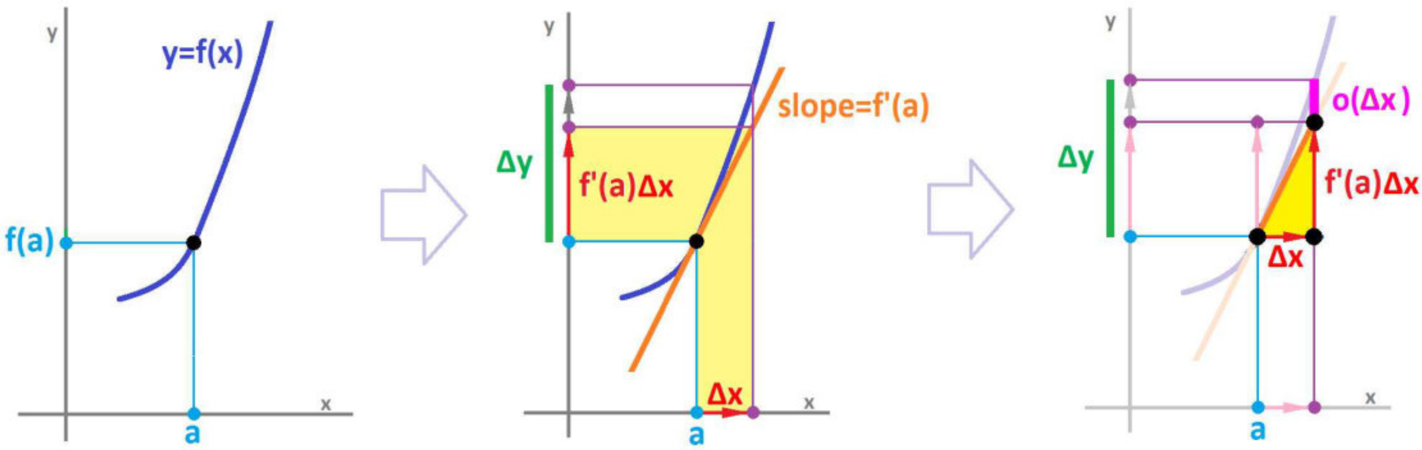
or

$$\lim_{\Delta x \to 0} \frac{\Delta y - f'(a)\Delta x}{\Delta x} = 0\,.$$

In other words,

$$\Delta y - f'(a)\Delta x << \Delta x \quad \text{as} \quad \Delta x \to 0\,.$$

We can see how these quantities interact on the graph:



We conclude the following:

Theorem 6.1.3: Derivative via Little o

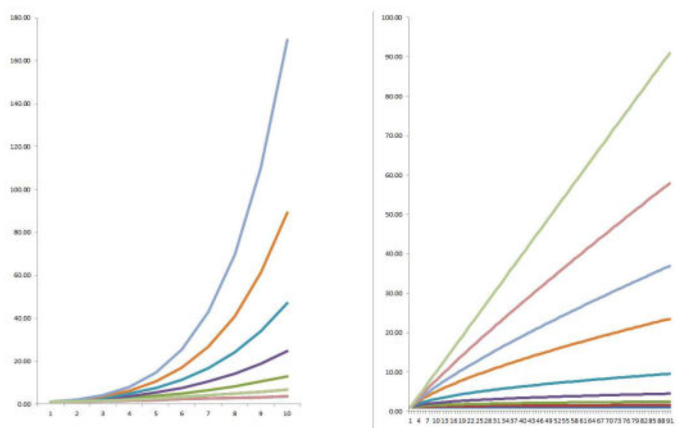
A function f is differentiable at $x=a$ if and only if:

$$\boxed{\Delta y = f'(a)\Delta x + o(\Delta x)}$$

The method presented above allows us to justify the following hierarchy at $+\infty$:

$$\boxed{... >> x^n >> ... >> x^2 >> x >> \sqrt{x} >> \sqrt[3]{x} >> ... >> 1}$$

Below, we see powers below 1 on the right and above 1 on the left:



The right is to be appended at the bottom of the left.

Among polynomials, the *degree* plays the role of the order:

Theorem 6.1.4: Little o for Polynomials

A polynomial has a higher order than another if and only if its degree is higher.

In other words, for any two polynomials P and Q, we have:

$$P = o(Q) \iff \deg P < \deg Q\,.$$

And, furthermore,

$$P \sim Q \iff \deg P = \deg Q\,.$$

Exercise 6.1.5

Prove the theorem.

We have, therefore, expanded our hierarchy:

$$\boxed{...>> n\text{th degree polynomial} >> ... >> \text{2nd degree polynomial} >> \text{1st degree polynomial}}$$

Then, what is the relation to the list of the powers that we started with? The meaning of each item in the sequence is the simplest representative, x^n, of a whole class of functions: the polynomials of nth degree.

Exercise 6.1.6

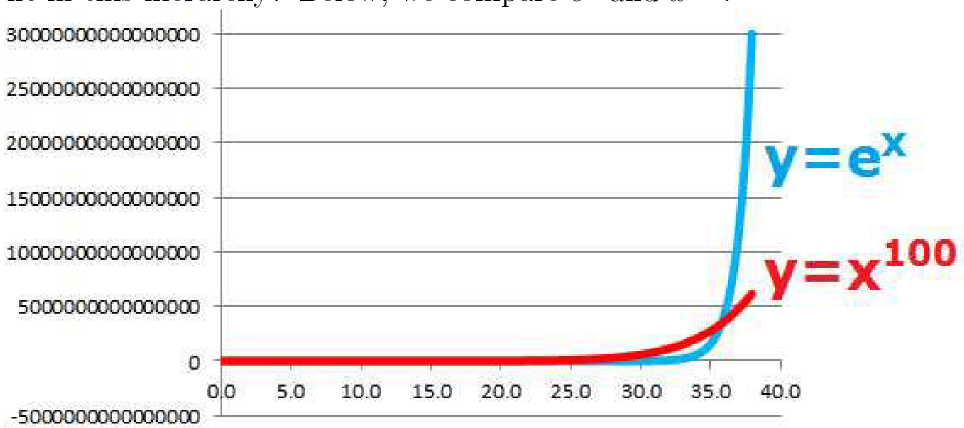
Can you further enlarge the classes of functions in this sequence? Hint: add $\sqrt{x}$.

Exercise 6.1.7

Suggest a "Little o Theorem For Rational Functions". Prove it.

Example 6.1.8: order at e^x

Where does e^x fit in this hierarchy? Below, we compare e^x and x^{100}:



The former seems stronger but, unfortunately, the methods of dividing by the higher power doesn't apply here. We will be looking for another method.

Next, we consider how to compare the *magnitudes of two functions in the vicinity of a point.* The idea of using limits to determine this relative order applies.

Example 6.1.9: order at 0

Consider the functions that form this fraction in the vicinity of 0:

$$\frac{3+x^{-2}-7x^{-3}}{2+10x^{-3}} \to -\frac{7}{10} \quad \text{as } x \to 0\,.$$

We conclude that these two functions have the "same order" at 0. However, below the numerator is of higher order at 0 than the denominator:

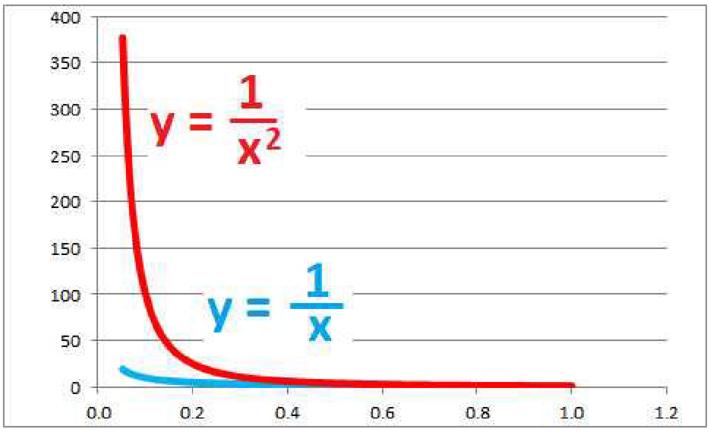
$$\frac{3+x^{-2}-7x^{-4}}{2+10x^{-3}} \to \infty \quad \text{as } x \to 0\,.$$

We are thus able to justify the following hierarchy at 0:

$$\boxed{... >> \frac{1}{x^n} >> ... >> \frac{1}{x^2} >> \frac{1}{x} >> \frac{1}{\sqrt{x}} >> \frac{1}{\sqrt[3]{x}} >> ... >> 1}$$

Example 6.1.10: reciprocals

Below, we compare $1/x$ and $1/x^2$:



However, the method fails for other types of functions; for example, where does $\ln x$ fit in this hierarchy? We will be looking for another method.

Let's observe that the order of a function (at infinity) is determined by its change, i.e., its *difference*. So, here is an idea:

▶ To compare two functions, we compare their differences instead.

In other words, we compute:

$$\lim_{x\to\infty}\frac{\Delta f}{\Delta g}(x)\,.$$

What about the *difference quotients*? The denominators cancel:

$$\frac{\frac{\Delta f}{\Delta x}}{\frac{\Delta g}{\Delta x}}=\frac{\Delta f}{\Delta g}\,.$$

So, this idea is just as valid:

▶ To compare two functions, we compare their difference quotients instead.

In other words, we compute:

$$\lim_{x\to\infty}\frac{\frac{\Delta f}{\Delta x}(x)}{\frac{\Delta g}{\Delta x}(x)}\,.$$

Finally, the order of a function is determined by its rate of growth, i.e., its *derivative*. We take the idea to the next level:

▶ To compare two functions, we compare their derivatives instead.

After all, the derivative *is* a limit, by definition. We have computed so many derivatives by now that we can use the results to evaluate some other limits. In other words, we compute:

$$\lim_{x\to\infty}\frac{\frac{df}{dx}(x)}{\frac{dg}{dx}(x)}\,.$$

Whether x is approaching infinite or a number, the approach is the same:

Theorem 6.1.11: L'Hopital's Rule

Suppose functions f and g are continuously differentiable. Then, for either of the two types of limits:

$$x\to\pm\infty \quad\text{and}\quad x\to a^{\pm}\,,$$

we have:

$$\boxed{\lim_{\square} \frac{f(x)}{g(x)} = \lim_{\square} \frac{f'(x)}{g'(x)}}$$

whenever the latter limit exists (as a number or infinity), and provided that the left-hand side is an indeterminate expression:

$$\lim_{\square} f(x) = \lim_{\square} g(x) = 0\,,$$

or

$$\lim_{\square} f(x) = \lim_{\square} g(x) = \infty\,.$$

Warning!

L'Hopital's Rule is *not* the Quotient Rule (of limits or of derivatives). The former is, in fact, a way to resolve indeterminacy that precludes application of the latter.

Proof.

We can justify this idea for the following simplified case. Let c be the point we are interested in and suppose that we have two functions f and g with:

$$f(c) = g(c) = 0\,.$$

We then can insert these expressions into their ratio:

$$\frac{f(x)}{g(x)} = \frac{f(x)-0}{g(x)-0} = \frac{f(x)-f(c)}{g(x)-g(c)} = \frac{\frac{f(x)-f(c)}{x-c}}{\frac{g(x)-g(c)}{x-c}}\,.$$

We also divided both by $g(x) - g(c)$, for $x \neq c$. We conclude from the *Quotient Rule* of limits:

$$\frac{f(x)}{g(x)} \to \frac{f'(c)}{g'(c)}\,,$$

provided that the derivatives exist and $g'(c) \neq 0$.

Example 6.1.12: fraction

Compute

$$\lim_{x\to\infty} \frac{x^2-3}{2x^2-x+1}\,.$$

It's indeterminate, ∞/∞. The known method is to divide by the highest power. Instead, we apply *L'Hopital's Rule*:

$$= \lim_{x\to\infty} \frac{2x}{4x-1}\,.$$

But it's still indeterminate! That's a good news: We can apply L'Hopital's Rule again! This is the result:

$$= \lim_{x\to\infty} \frac{2}{4} = \frac{1}{2}\,.$$

Warning!

Before you apply L'Hopital's Rule, verify that this is indeed an indeterminate expression.

Example 6.1.13: simplification

Compute

$$\begin{aligned} &\lim_{x\to 1}\frac{\ln x}{x-1} && \text{Resulting in } \frac{\infty}{\infty}\text{? Apply LR.}\\ &=\lim_{x\to 1}\frac{\frac{1}{x}}{1}\\ &=\lim_{x\to 1}\frac{1}{x}\\ &=1\,. \end{aligned}$$

Example 6.1.14: exponential function

Compute

$$\begin{aligned} &\lim_{x\to\infty}\frac{e^x}{x^2} && \text{Resulting in } \frac{\infty}{\infty}\text{? Apply LR.}\\ &=\lim_{x\to\infty}\frac{e^x}{2x} && \text{Resulting in } \frac{\infty}{\infty}\text{? Apply LR.}\\ &=\lim_{x\to\infty}\frac{e^x}{2}\\ &=\infty\,. \end{aligned}$$

Therefore,

$$e^x >> x^2\,.$$

In fact, no matter how high the degree is, a polynomial comes to zero after a sufficient number of differentiations. Meanwhile, nothing happens to the exponent:

$$\lim_{x\to\infty}\frac{e^x}{x^n}=\infty\,.$$

Therefore, we have our hierarchy appended:

$$\boxed{e^x >> ... >> x^n >> ... >> x^2 >> x >> \sqrt{x} >> \sqrt[3]{x} >> ... >> 1}$$

Example 6.1.15: larger than exponential

Is there are a larger one than e^x? Yes, e^{2x}. Larger than that? Yes, e^{x^2}. And so on.

Exercise 6.1.16

Add more to the list. On the right too.

Exercise 6.1.17

Is there a largest function?

Exercise 6.1.18

Is it possible to insert functions in this list?

$$x^2 >> f >> x \quad\text{and}\quad x >> g >> \sqrt{x}$$

What about other indeterminate expressions? Below are the possible types of limits that correspond to algebraic operations:

- products:
$$0 \cdot \infty$$
- differences:
$$\infty - \infty$$
- powers:
$$0^0,\ \infty^0,\ 1^\infty$$

Warning!

An indeterminate expression on its own is meaningless. When it comes from a limit to be evaluated, its meaning is DEAD END.

The idea is to *convert them to fractions.*

Example 6.1.19: products

Evaluate:
$$\lim_{x\to 0^+} x\ln x = ?$$
How do we apply L'Hopital's Rule here? We convert to a fraction by dividing by the reciprocal:
$$\begin{aligned}
&= \lim_{x\to 0^+} \frac{\ln x}{\frac{1}{x}} && \text{Resulting in } \frac{\infty}{\infty}\text{? Apply LR.}\\
&= \lim_{x\to 0^+} \frac{\frac{1}{x}}{-\frac{1}{x^2}} && \text{Simplify!}\\
&= \lim_{x\to 0^+} -x = 0\,.
\end{aligned}$$

Example 6.1.20: exponents

Evaluate
$$\lim_{x\to\infty} x^{\frac{1}{x}} = ?$$
How do we convert to a fraction? Use the logarithm!
$$\ln x^{\frac{1}{x}} = \frac{1}{x}\ln x\,.$$
Then, we use the fact that the logarithm is continuous (the "magic words"):
$$\begin{aligned}
\ln\left(\lim_{x\to\infty} x^{\frac{1}{x}}\right) &= \lim_{x\to\infty}\left(\ln x^{\frac{1}{x}}\right)\\
&= \lim_{x\to\infty}\left(\frac{1}{x}\ln x\right) && \text{This is a fraction!}\\
&= \lim_{x\to\infty}\frac{\ln x}{x} && \text{Resulting in } \frac{\infty}{\infty}.\\
&\overset{\text{LR}}{=\!=} \lim_{x\to\infty}\frac{\frac{1}{x}}{1} = 0\,.
\end{aligned}$$
Therefore,
$$\lim_{x\to\infty} x^{\frac{1}{x}} = e^0 = 1\,.$$

Exercise 6.1.21

What property of limits did we use at the end?

Exercise 6.1.22

Prove $x >> \ln x >> \sqrt{x}$ at ∞.

Exercise 6.1.23

Compare $\ln x$ and $\dfrac{1}{x^n}$ at 0.

Example 6.1.24: misapplication

The condition of theorem that requires the ratio to be an indeterminate expression must be verified.

What can happen otherwise is illustrated below. If we apply *L'Hopital's Rule*, we get the following:

$$\lim_{x\to+\infty}\frac{1+\frac{1}{x}}{1+\frac{1}{x^2}} \overset{\text{LR?}}{=\!=} \lim_{x\to+\infty}\frac{\left(1+\frac{1}{x}\right)'}{\left(1+\frac{1}{x^2}\right)'} = \lim_{x\to+\infty}\frac{-\frac{1}{x^2}}{-\frac{2}{x^3}} = \frac{1}{2}\lim_{x\to+\infty} x = +\infty\,.$$

But the original limit doesn't produce an indeterminate expression! This makes the *Quotient Rule* for limits applicable:

$$\lim_{x\to+\infty}\frac{1+\frac{1}{x}}{1+\frac{1}{x^2}} \overset{\text{QR}}{=\!=} \frac{\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)}{\lim_{x\to+\infty}\left(1+\frac{1}{x^2}\right)} = \frac{1}{1} = 1\,.$$

A mismatch! L'Hopital's Rule is inapplicable because the limits of the numerator and the denominator of the original fraction *exist*! Here is another example:

$$\lim_{x\to 1}\frac{x^2}{x} \overset{\text{LR?}}{=\!=} \lim_{x\to 1}\frac{2x}{1} = 2\,.$$

Exercise 6.1.25

Describe this class of functions: $o(1)$.

In summary, in computing the limit of a fraction, it is only possible to follow *one* of these two routes:

$$\begin{array}{ccccc} & & \boxed{\text{Quotient Rule}} & \to & \dfrac{\lim f}{\lim g} \\ & \nearrow & & & \\ \lim \dfrac{f}{g} & & & & \\ & \searrow & & & \\ & & \boxed{\text{L'Hopital's Rule}} & \to & \dfrac{\lim f'}{\lim g'} \end{array}$$

6.2. *Linear approximations*

It is a reasonable strategy to answer a question that you don't know how to answer by answering instead another one – close to the original – with a known answer. For example:

- What is the square root of 4.1? I know that $\sqrt{4} = 2$, so I'll say that it's about 2.

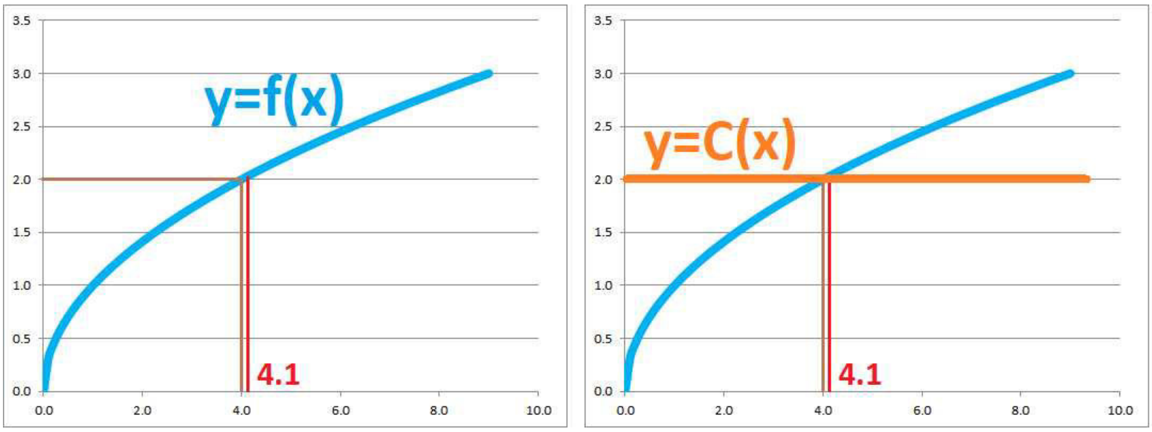
- What is the square root of 4.3? It's about 2.
- What is the square root of 3.9? It's about 2.

And so on.

These are all reasonable estimates, but they are all the same:

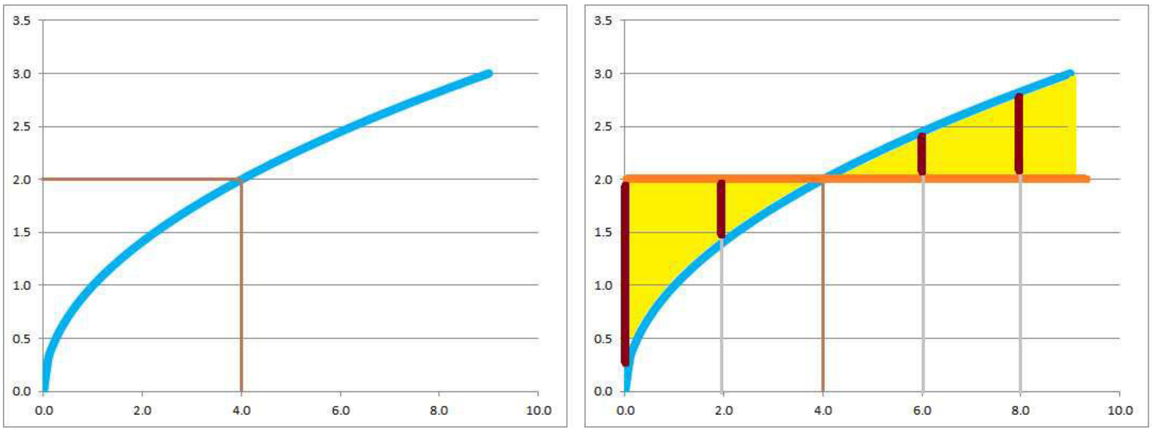
x	3.8	3.9	4.0	4.1	4.2
y	2	2	2	2	2

Our interpretation of this observation is that we speak of approximating the function $f(x) = \sqrt{x}$ by a *constant* function $C(x) = 2$:



It is a valid but crude approximation!

An approximation of a function is just another function. To make sense of it, let's consider the *error of an approximation* as the difference between the two functions. For the case of a constant approximation, the errors are visible below as the segments in the yellow gap between the two graphs:



So, the error of an approximation is the following:

$$E(x) = |f(x) - C|\,.$$

But what C do we choose? Even though the choice is obvious, let's run through the argument anyway.

This is the requirement we put forward:

▶ The error diminishes as x is getting closer to a.

In other words, we would like the following to be satisfied:

$$E(x) = |f(x) - C| \to 0 \quad \text{as} \quad x \to a\,.$$

Therefore, we require:

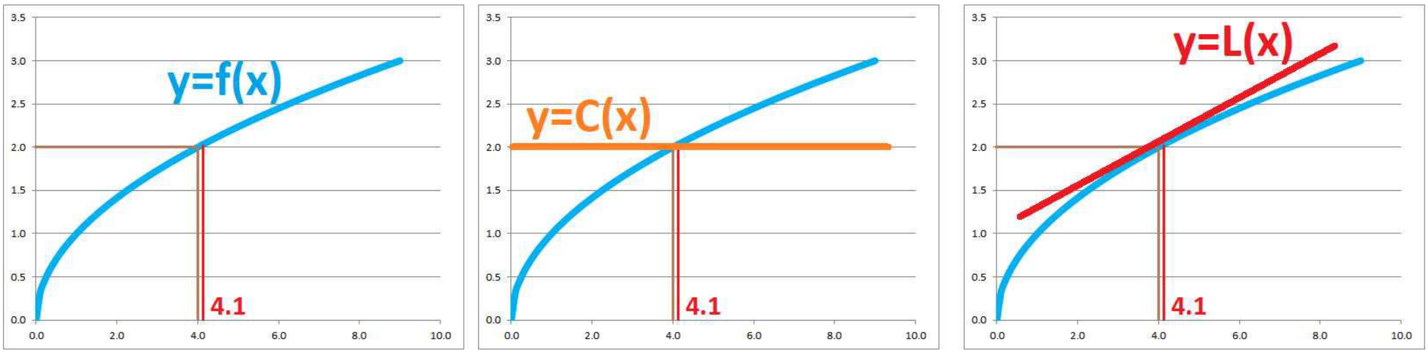
$$f(x) \to C \quad \text{as} \quad x \to a\,.$$

We discover that it suffices to ask f to be *continuous* at a and to choose $C = f(a)$! We are done.

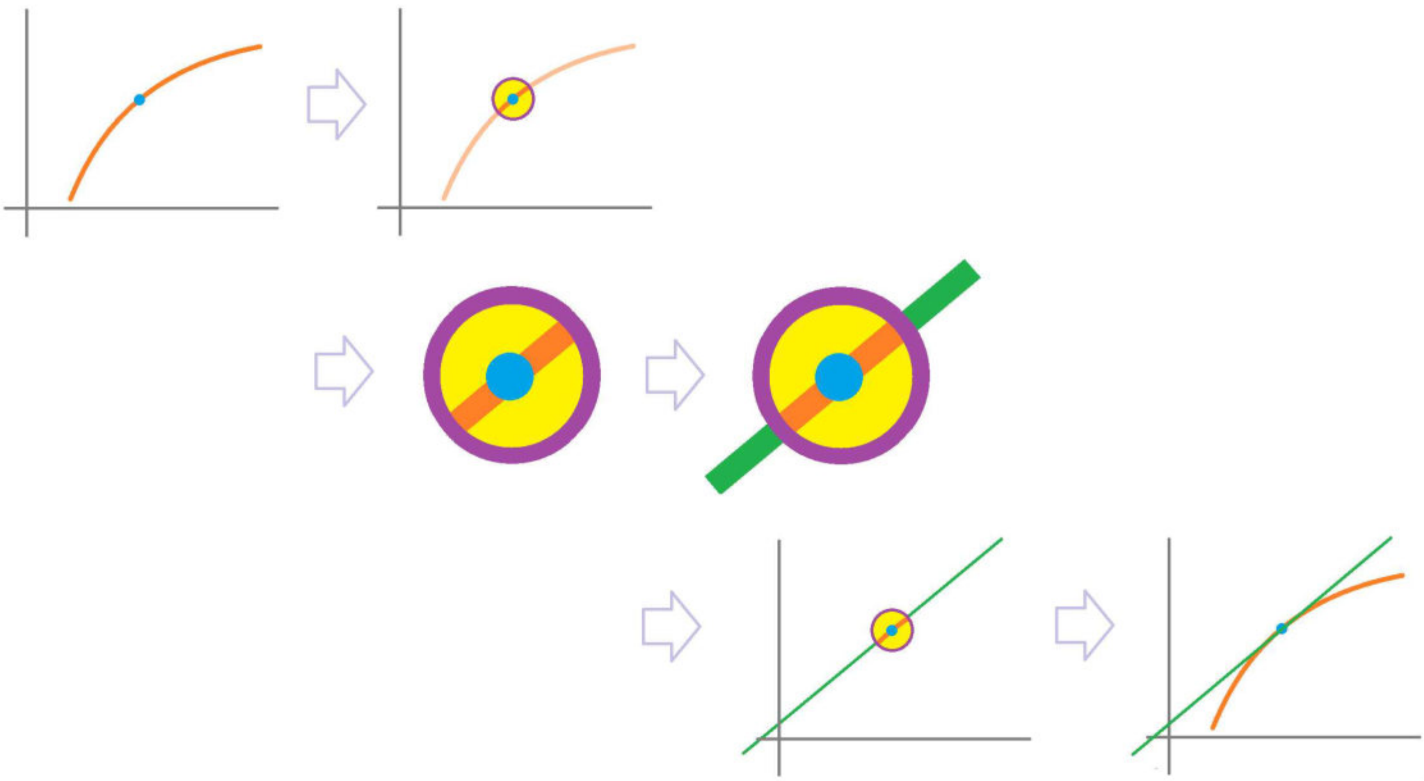
Let's take this to the next level:

▶ Can we do better than the horizontal line?

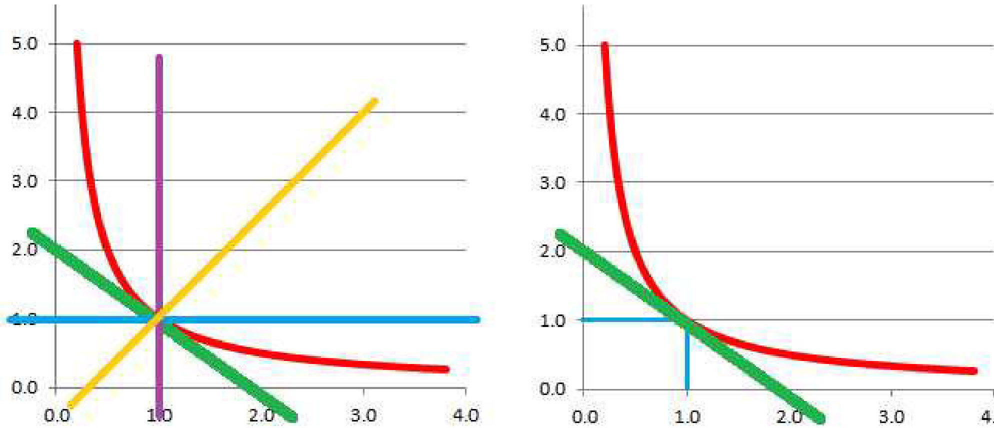
Definitely:



We already know that the tangent line "approximates" the graph of a function; in fact, it was designed to do that exactly! The idea was that when you zoom in on the point, the tangent line will merge with the graph:



However, there are *many* straight lines that can be used to approximate, including the horizontal line:



What is so special about the function represented by the tangent line?

We are about to start using more precise language. Recall the *point-slope form* (seen in Volume 1, Chapter

1PC-2) of the line through (x_0, y_0) with slope m as a relation:

$$y - y_0 = m(x - x_0)\,.$$

Specifically, we require the candidates to at least pass though the point of interest:

$$x_0 = a,\ y_0 = f(a)\,.$$

Therefore, we have:

$$y - f(a) = m(x - a)\,.$$

At this point, we make an important step and stop looking at this as an equation of a line but as a formula for a new *function*:

$$\boxed{L(x) = f(a) + m(x - a)}$$

It is a linear function: The power of x is 1, while the rest of the parameters are constant. That's why it is called a *linear approximation* of f at $x = a$.

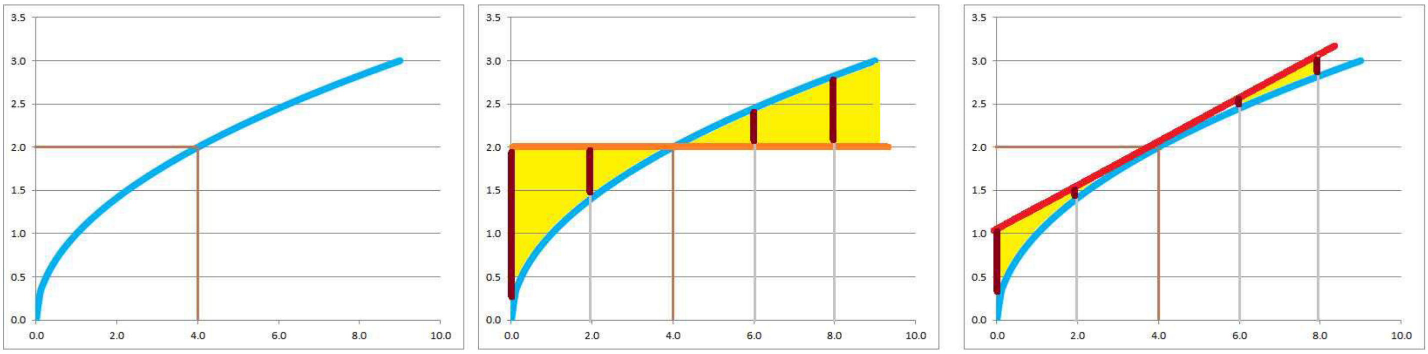
Now, *algebra*. Once again, let's look at the error, i.e., the difference between the two functions:

$$E(x) = |f(x) - L(x)|\,.$$

All of the three functions are *continuous* at a! Therefore, the error will diminish as x is gets closer to a:

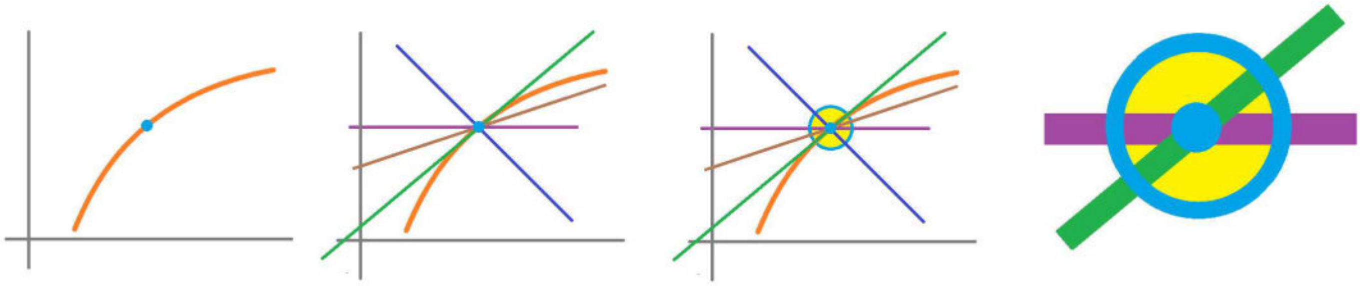
$$E(x) = |f(x) - L(x)| \to 0 \text{ as } x \to a\,.$$

In other words, every linear function $y = L(x)$ defined as above will pass the test that we used to choose the best constant approximation:



That's not good enough anymore!

Let's look at these linear functions, $y = L(x)$, as a group. When you zoom in on the point, we realize the *angle* between the lines is preserved under the magnification:



How do we choose the "best" approximation? The only difference among them is how fast the error is diminishing. This is the plan:

- ▶ We will choose the linear function with the error converging to 0 the fastest possible.

Specifically, we *compare* the convergence of the error with that of the distance function:

$$E(x) = |f(x) - L(x)| \to 0 \ \text{ vs. } \ |\Delta x| = |x - a| \to 0\,.$$

How do we compare the "speed of convergence"?

We know the answer from our study of the magnitudes of function earlier in this chapter: Look at the convergence of their ratio! Then, $E(x)$ converges faster than $|\Delta x|$ if the limit of this ratio is zero; i.e.,

$$\frac{f(x)-L(x)}{x-a} \to 0 \text{ as } x \to a\,.$$

In other words, we have:

$$f(x)-L(x) = o(x-a)\,.$$

Surprisingly, this general condition is sufficient to make the slope m of L specific!

Theorem 6.2.1: Best Linear Approximation

Suppose f is differentiable at $x=a$ and

$$L(x) = f(a) + m(x-a)$$

is any of its linear approximations. Then,

$$\lim_{x\to a} \frac{f(x)-L(x)}{x-a} = 0 \iff m = f'(a)\,.$$

Proof.

As $x \to a$, we have:

$$\begin{array}{rcccc} \dfrac{f(x)-(f(a)+m(x-a))}{x-a} & = & \dfrac{f(x)-f(a)}{x-a} & -m \\ & \to & f'(a) & -m\,. \end{array}$$

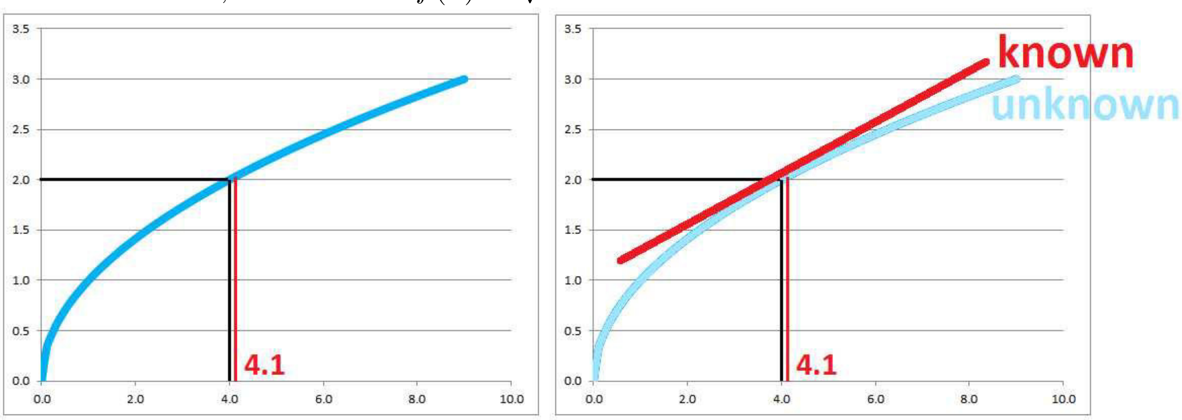
Warning!

The best linear approximation is often simply called the "linear approximation".

Example 6.2.2: root

Let's approximate $\sqrt{4.1}$.

We can't compute $\sqrt{x}$ by hand. In fact, the only meaning of $x=\sqrt{4.1}$ is that it is such a number that $x^2 = 4.1$. In that sense, the function $f(x)=\sqrt{x}$ is *unknown*:



Just a few exceptions are $f(4)=\sqrt{4}=2$, $f(9)=\sqrt{9}=3$, etc. We will use these points as "anchors". Then, we can use the *constant approximation* and declare:

$$\sqrt{4.1} \approx 2\,.$$

We are done if the quality of the approximation is sufficient.

Just as the function $f(x) = \sqrt{x}$ is unknown, so is its derivative:

$$f'(x) = \frac{1}{2\sqrt{x}}\,.$$

But, just as f is known at $x = 4$, so is f':

$$f'(4) = \frac{1}{2\sqrt{4}} = \frac{1}{4}\,.$$

What can we do with this information?

The best linear approximation of f is *known* in the sense that, as a linear function, it can be computed by hand:

$$\begin{array}{lll} L(x) & = f(a) & +f'(a)(x-a) \\ & = 2 & +\frac{1}{4}(x-4)\,. \end{array}$$

This function is a *replacement* for $f(x) = \sqrt{x}$ in the vicinity of the "anchor point" $x = 4$.

Finally, we compute the approximation *by hand*:

$$\begin{aligned} L(4.1) &= 2 + \frac{1}{4}(4.1 - 4) \\ &= 2 + \frac{1}{4}\cdot .1 \\ &= 2 + 0.025 \\ &= 2.025\,. \end{aligned}$$

Our conclusion is that for the *linear approximation* we have:

$$\sqrt{4.1} \approx 2.025\,.$$

Exercise 6.2.3

Find the best constant and linear approximations of $f(x) = \sqrt{3.99}$.

In summary, the best linear approximation is given by the following formula:

$$\boxed{L(x) = f(a) + f'(a)(x-a)}$$

Warning!

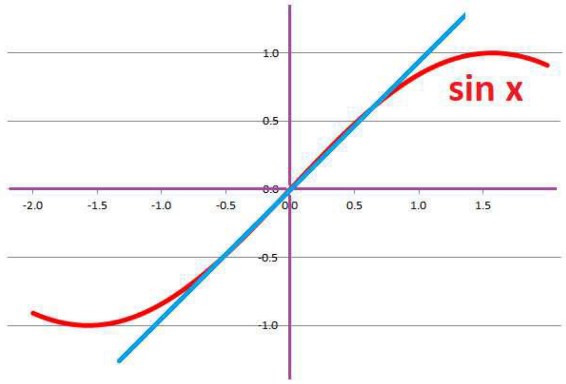
Don't plug in the formula for the derivative into the formula for the best linear approximation.

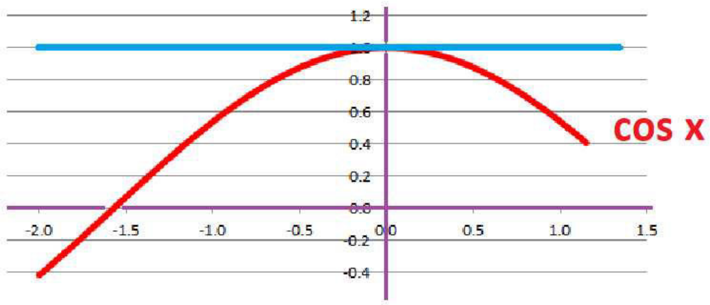
Exercise 6.2.4

Find the best linear approximation of $f(x) = x^{1/3}$ at $a = 1$. Use it to estimate $1.1^{1/3}$.

Example 6.2.5: sine and cosine

Finding the best linear approximation amounts to finding the tangent line at the point. Some are already known:





Let's redo them:

functions:	$y = \sin x$		$y = \cos x$	
evaluated at 0 :	$\sin x\big\|_{x=0}$	$= \sin 0 = 0$	$\cos x\big\|_{x=0}$	$= \cos 0 = 1$
derivatives:	$\big(\sin x\big)'$	$= \cos x$	$\big(\cos x\big)'$	$= -\sin x$
evaluated at 0 :	$\big(\sin x\big)'\big\|_{x=0}$	$= \cos 0 = 1$	$\big(\cos x\big)'\big\|_{x=0}$	$= -\sin 0 = 0$
tangents:	$y - 0 = 1 \cdot (x - 0)$		$y - 1 = 0 \cdot (x - 0)$	
linear approximations:	$L(x) = x$		$l(x) = 1$	

Then, we have

$$\sin .2 \approx .2, \ \sin -.01 \approx -.01, \ \text{etc.}$$

and

$$\cos .2 \approx 1, \ \cos -.01 \approx 1, \ \text{etc.}$$

Exercise 6.2.6

What are the best constant approximations of the sine and the cosine?

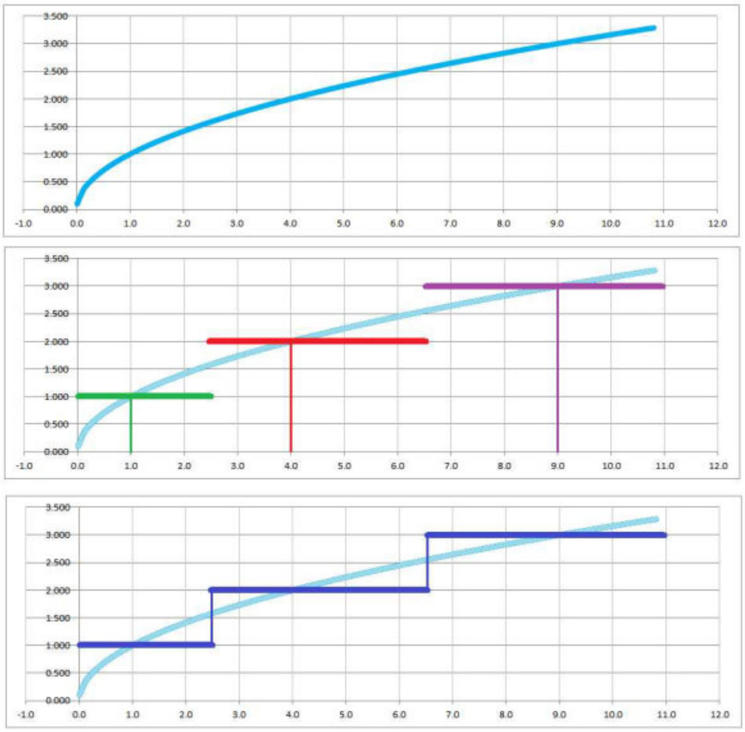
Exercise 6.2.7

Use the best linear approximation of $f(x) = \sqrt{\sin x}$ to estimate $\sqrt{\sin \pi/2}$.

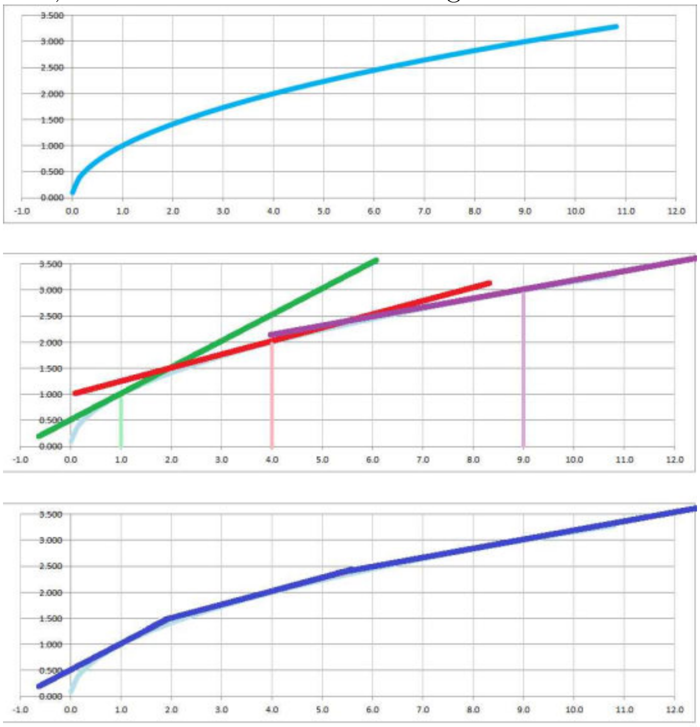
Example 6.2.8: piecewise-defined

If we were to design a *calculator*, we'd need enough "anchor" points in order to produce an approximation of the whole, unknown, function. For $f(x) = \sqrt{x}$, as long as the number is "close" to 4, we use it or we look for another number with a known square root. For example, $\sqrt{.99} \approx \sqrt{1} = 1$, $\sqrt{10} \approx \sqrt{9} = 3$, and so on.

For the constant approximation, we switch half-way between the points:



For the linear approximation, we switch wherever the tangents intersect:



Exercise 6.2.9

What makes approximating the function around 0 seem to be poor?

Replacing a function with its linear approximation is called *linearization*. Linearizations make a lot of things much simpler.

Example 6.2.10: differentiation and integration

The sine and cosine are related to each other via differentiation and integration. Are their best linear approximations related in that manner? Yes, but not exactly.

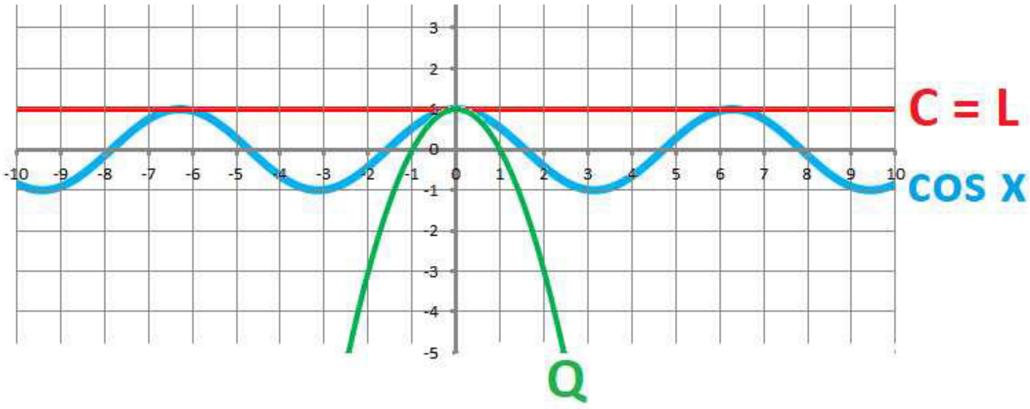
Differentiation:

$$\begin{array}{llcl} \text{functions:} & \sin x & \to \dfrac{d}{dx} \to & \cos x \\ \text{approximations:} & x & \to \dfrac{d}{dx} \to & 1 \\ \text{polynomials:} & \text{linear} & \to \dfrac{d}{dx} \to & \text{constant} \end{array}$$

Integration:

$$\begin{array}{llcl} \text{functions:} & \sin x & \to \displaystyle\int \to & -\cos x + C \\ \text{approximations:} & x & \to \displaystyle\int \to & \dfrac{1}{2}x^2 + C \\ \text{polynomials:} & \text{linear} & \to \displaystyle\int \to & \text{quadratic} \end{array}$$

The antiderivative of $\sin x$ is the *quadratic* approximation of $-\cos x$:



Linearizations may help explaining the ideas of continuity and differentiability.

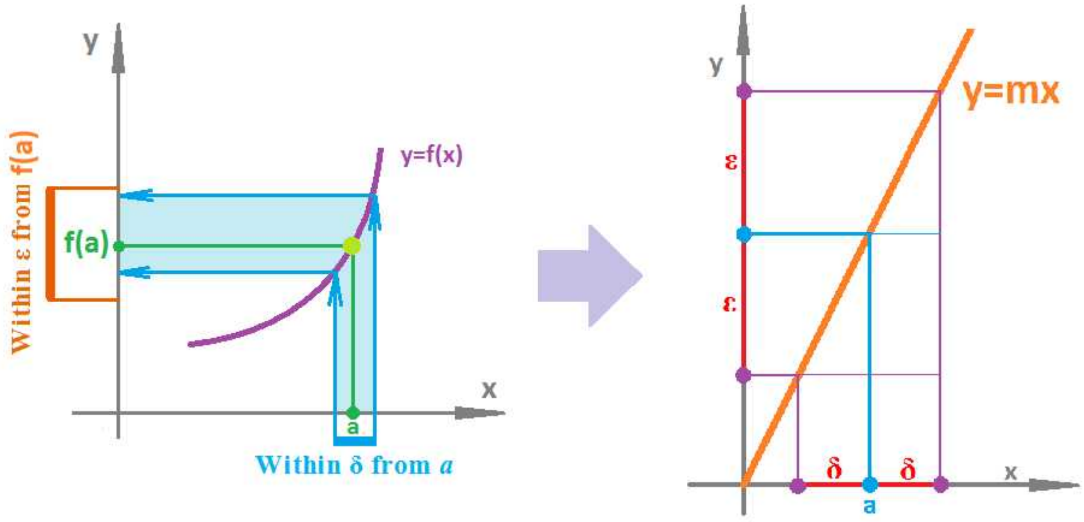
Example 6.2.11: continuity

All linear functions, $L(x) = mx + b,\ m \neq 0$, are, of course, *continuous*. Furthermore, the relation, in the definition of continuity (Chapter 2), between ε and δ is very simple: To ensure

$$|x - a| < \delta \implies |L(x) - L(a)| < \varepsilon\,,$$

we simply choose:

$$\delta = \frac{1}{m}\varepsilon\,.$$



We now interpret these two, as before:

- $\delta = \Delta x$ is the accuracy of the measurement of x.
- $\varepsilon = \Delta y$ is the accuracy of the indirect evaluation of $y = L(x)$.

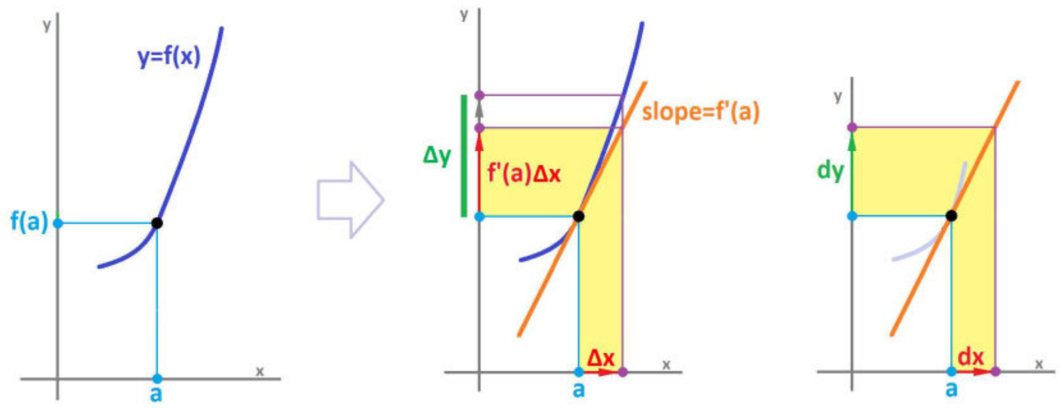
Then, of course, we have:

$$\Delta y = m\Delta y \quad \text{or} \quad \Delta y = L'(a)\Delta y\,.$$

We can have a similar, but this time approximate, analysis for any function$f$:

▶ We can make $\Delta y = f(a + \Delta x) - f(a)$ as small as we like by choosing Δx small enough.

How do we find this Δx? We *linearize*:



The function is replaced with its best linear approximation at $x = a$ and the analysis in the last example is applied to this function:

$$f'(a) = \left.\frac{dy}{dx}\right|_{x=a} \approx \frac{\Delta y}{\Delta x}\,.$$

Then, the change of the output variable is approximately proportional to the change of the input variable:

$$\boxed{\Delta y \approx f'(a)\,\Delta x}$$

More precisely, these two functions are of the same order as explained in the last section:

$$\Delta y \sim f'(a)\,\Delta x\,.$$

How well this approximation works is discussed in the next section.

This idea can also be expressed via the *differentials*:

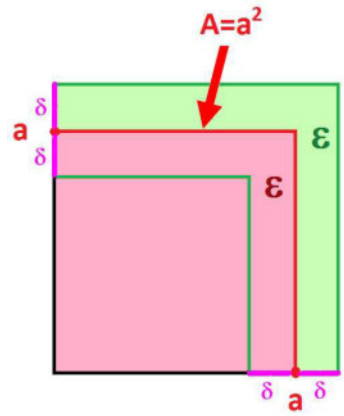
$$\boxed{dy = f'(a)\,dx}$$

The equation is, in fact, their definition.

Example 6.2.12: error estimation

Let's revisit the example of evaluating the area $A = f(x) = x^2$ of square tiles when their dimensions are close to 10×10. The area is supposed to be 100, but how close to the truth is this number? The question may be important for passing a quality inspection.

We look at how the error of the *computation* of the area depends on the error of the *measurement* of the side x:



Suppose the desired accuracy of A is $\Delta A = 5$, what should be the accuracy Δx of x? By brute force, we discovered that $\Delta x = .2$ is appropriate:

$$\begin{aligned} A &= (10 \pm .2)^2 \\ &= 10^2 \pm 2 \cdot 10 \cdot .2 + .2^2 \\ &\text{or } 100.04 \pm 4\,. \end{aligned}$$

This time, instead, we get a quick "ballpark" figure by using the linearization of $f(x) = x^2$ at $x = 10$. We find the derivative, $2x$, at $x = 10$:

$$f'(10) = 2 \cdot 10 = 20\,.$$

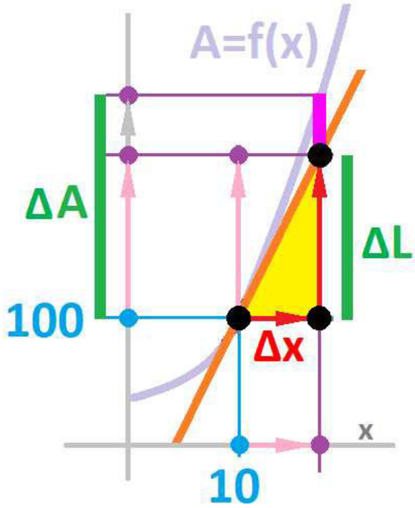
Then we use the formula for best linear approximation:

$$L(x) = 10 + 20(x - 10)\,.$$

Then

$$|\Delta A| = |A(10) - A(x)| \approx |L(10) - L(x)| = |20(x - 10)| = 20|\Delta x|\,.$$

We can see the relation below:



Therefore,

$$|\Delta x| \approx 5/20 = .25\,.$$

So, in order to achieve the accuracy of 4 square inches of the computation of the area of a 10×10 tile, one will need the 1/4-inch accuracy of the measurement of the side of the tile.

Exercise 6.2.13

What is the accuracy of the computation of the area if the accuracy for the side is .03?

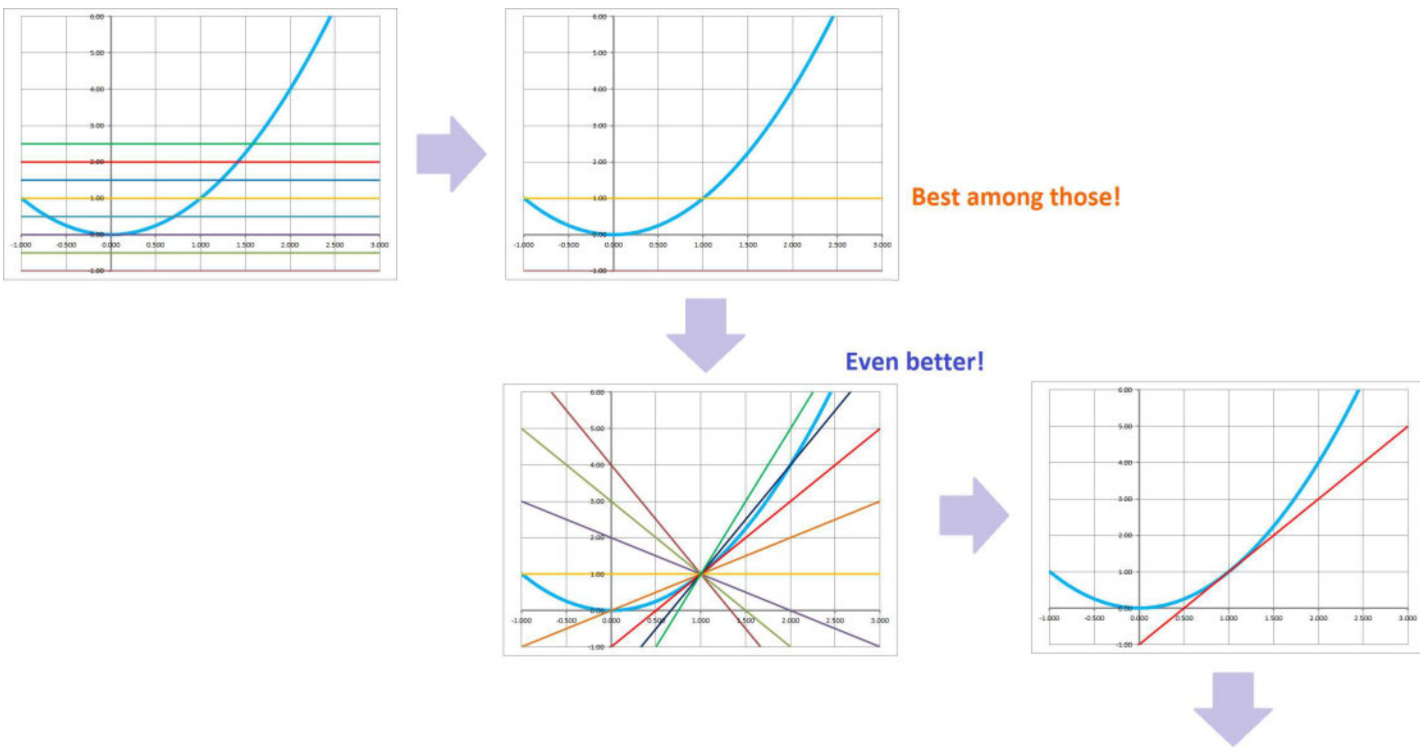
Exercise 6.2.14

What is the accuracy of the computation of the circumference (the length) of a circle if its radius is found to be 20 ± 1 inches? What about the area?

Exercise 6.2.15

What is the accuracy of the computation of the surface area of the Earth if the radius is found to be $6,360 \pm 30$ kilometers? What about the volume?

As a summary, below we illustrate how we attempt to approximate a function around the point $(1,1)$ with constant functions first: from those we choose the horizontal line through the point. This line then becomes one of many linear approximations of the curve that pass through the point: from those we choose the tangent line. The two steps are shown below:

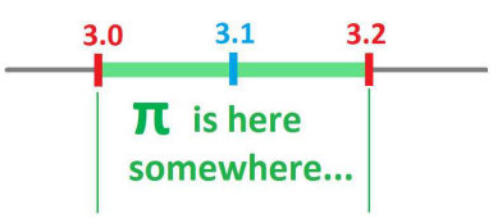


In Volume 3, Chapter 3IC-5, we will see that these are just the two first steps in the infinite sequence of approximations.

6.3. The accuracy of the best linear approximation

An "approximation" is meaningless if it comes as a single number.

We need to know more in order to make this number useful. For example, how close is 3.14 to π? Even more important is the question: how close is π to 3.14? The answer to the question will give π a definitive range of possible values:



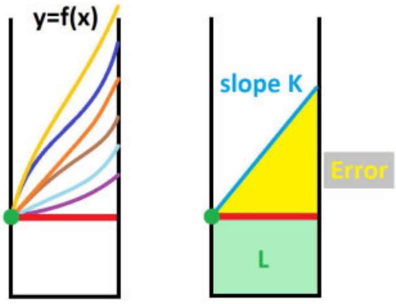
What's important is that π cannot be anywhere else!

Or, from the last section, how close is $\sqrt{4.1}$ to 2.025?

The answer will tell us where the truth lies, with *an absolute certainty if not absolute accuracy.*

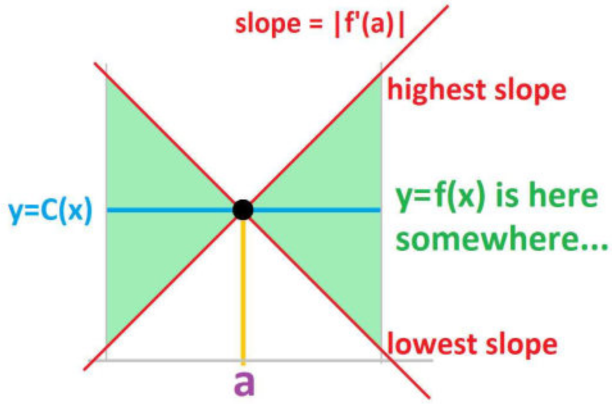
First, the *constant approximation.*

Even though the function $y = f(x)$ coincides with $y = C = f(a)$ at $x = a$, it may run away very fast and very far afterwards. How far? The only limit is the rate of growth of f, i.e., its derivative:



So, we can predict the behavior of f if we have *a priori* information about $|f'|$.

We will have a range of possible values for $f(x)$, for all x! The result is a *funnel* that contains the (unknown) graph of $y = f(x)$:



Indeed, we can conclude from the *Mean Value Theorem* that

$$|f(x) - C(x)| = |f(x) - f(a)| = |f'(t) \cdot (x - a)| = |f'(t)| \cdot |x - a| \le K|x - a| \,,$$

if we only know that $|f'(t)| \le K$ for all t between a and x.

Example 6.3.1: approximate square root

How close is $\sqrt{4.1}$ to 2? We need to find a bound for the derivative of $f(x) = \sqrt{x}$. Here we take into account that f' is decreasing and, therefore, its largest value is in the beginning:

$$f'(t) = \frac{1}{2\sqrt{t}} \le \frac{1}{2\sqrt{4}} \le .25 \,.$$

That could be K. Then we have bounds for the unknown $\sqrt{4.1}$:

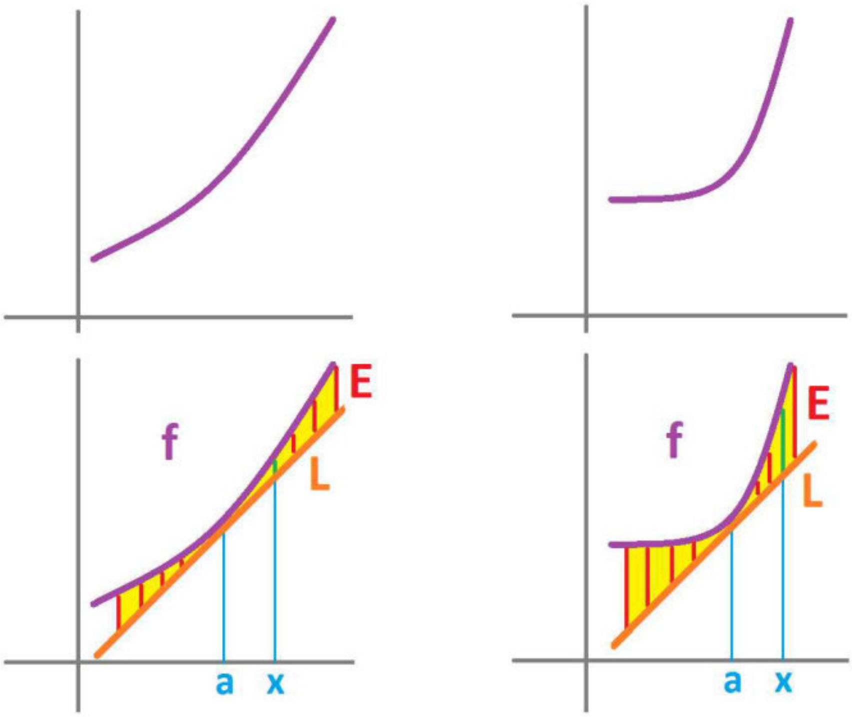
$$|\sqrt{4.1} - 2| \le .25 \cdot |4.1 - 4| = .025 \,.$$

In other words,

$$\sqrt{4.1} \quad \text{lies within} \quad [2 - .025, 2 + .025] = [1.975, 2.025] \,.$$

Next, the *linear approximation.*

The two graphs below have the same tangent line but the first one seems better approximated:



What causes the difference? The latter is more curved (concave)! The quantity that makes the slopes, i.e., the derivatives, change is the *second derivative*. We have an analogy:

- We know the accuracy of the best *constant* approximation if we have a bound on the magnitude of the *first* derivative.
- We know the accuracy of of the best *linear* approximation if we have a bound on the magnitude of the *second* derivative.

The latter takes the following form:

Theorem 6.3.2: Error Bound for Best Linear Approximation

Suppose f is twice differentiable at $x = a$ and $L(x) = f(a) + f'(a)(x - a)$ is its best linear approximation at a. Then we have:

$$\boxed{E(x) = |f(x) - L(x)| \le \frac{1}{2}K(x-a)^2}$$

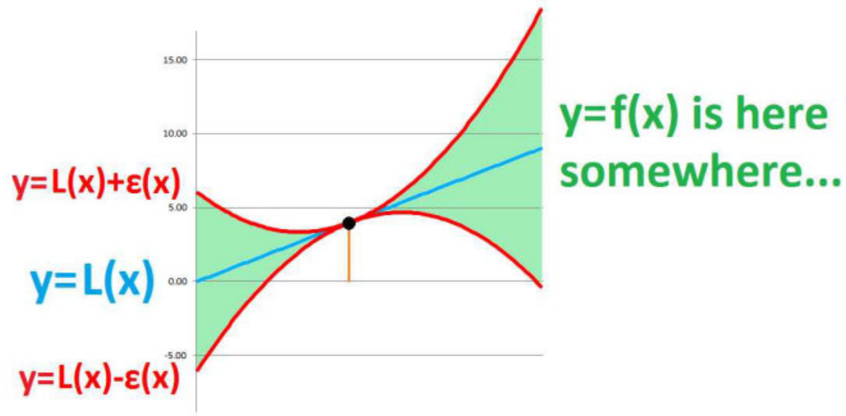
where K is a bound of the second derivative on the interval from a to x, i.e.,

$$|f''(t)| \le K \text{ for all } t \text{ in this interval.}$$

The theorem claims that

$$\boxed{E = o\big((x-a)^2\big)}$$

Once K is fixed, we have a range of possible values of the function $y = f(x)$! The result is, again, a *funnel* that contains the (unknown) graph of $y = f(x)$:



This time, the two edges of the funnel are parabolas:

$$y = L(x) \pm \frac{1}{2}K(x-a)^2 .$$

Therefore, the value of K makes the funnel proportionally wider or narrower.

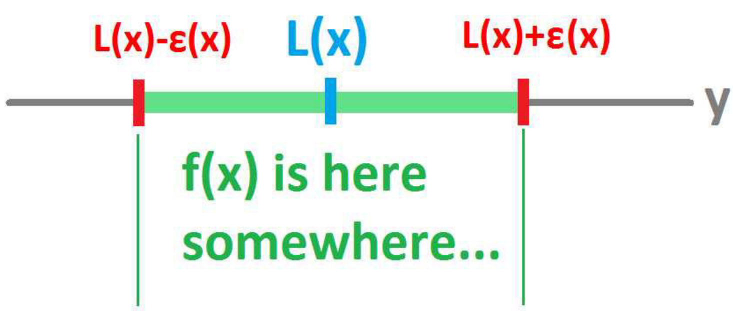
The practical meaning of the theorem is the following: For each x, this error bound,

$$\varepsilon(x) = \frac{1}{2}K(x-a)^2 ,$$

is the *accuracy of the approximation* in the sense that the interval located on the y-axis

$$\big[L(x) - \varepsilon(x), L(x) - \varepsilon(x)\big]$$

contains the true number, $f(x)$. We see this below:



This interval is a vertical cross-section of the funnel.

Example 6.3.3: root

We continue with the last example:

$$\sqrt{4.1} \approx L(4.1) = 2.025 .$$

Once again, the answer is unsatisfactory because it doesn't really tell us anything about the true value of $\sqrt{4.1}$. In fact, $\sqrt{4.1} \neq 2.025$!

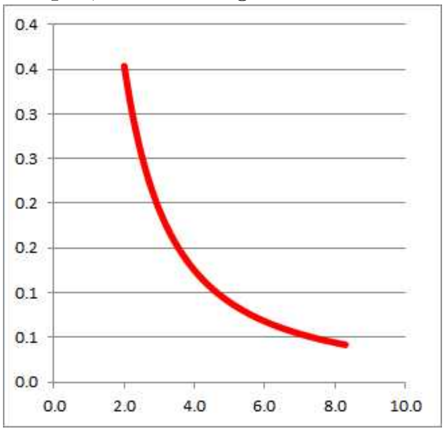
Let's apply the theorem. First, we find the second derivative:

$$\begin{aligned}
f(x) &= \sqrt{x} \implies f'(x) = \frac{1}{2\sqrt{x}} \implies \\
f''(x) &= \left(\frac{1}{2\sqrt{x}}\right)' = \left(\frac{1}{2}x^{-1/2}\right)' && \text{According to Power Formula.} \\
&= \frac{1}{2}\left(-\frac{1}{2}\right)x^{-1/2-1} \\
&= -\frac{1}{4}x^{-3/2} .
\end{aligned}$$

So, we need a bound for this function:

$$|f''(x)| = \frac{1}{4}\left|x^{-3/2}\right|$$

over the interval $[4, 4.1]$. This is a simple, decreasing function:



Therefore,

$$|f''(x)| \leq |f''(4)| = \frac{1}{4}\left|4^{-3/2}\right| = \frac{1}{32} = 0.03125\,.$$

This is our best choice for K! According to the theorem, our conclusion is that the error of the approximation cannot be larger than the following:

$$E(x) = |f(x) - L(x)| \leq \frac{1}{2}0.03125 \cdot (x-4)^2\,.$$

Specifically,

$$E(4.1) = \left|\sqrt{4.1} - L(4.1)\right| \leq \frac{1}{2}0.03125 \cdot (4.1-4)^2\,,$$

or

$$\left|\sqrt{4.1} - 2.025\right| \leq 0.00015625\,.$$

Therefore, we conclude:

- ▶ $\sqrt{4.1}$ is within $\varepsilon = 0.00015625$ from 2.025.

In other words, we have:

$$2.025 - 0.00015625 \leq \sqrt{4.1} \leq 2.025 + 0.00015625\,,$$

or

$$2.02484375 \leq \sqrt{4.1} \leq 2.02515625\,.$$

We have found an interval that is *guaranteed* to contain the number we are looking for!

Example 6.3.4: sin

Approximate $\sin .01$. Note that the constant approximation is $\sin .01 \approx \sin 0 = 0$. So, we have $a = 0$. Now we compute:

$$\begin{array}{llll} f(x) = \sin x & & & \\ f'(x) = \cos x & \Longrightarrow & L(x) = 0 + \cos x\Big|_{x=0}(x-0) & \Longrightarrow \quad L(x) = x \\ f''(x) = -\sin x & \Longrightarrow & |f''(x)| = \sin x & \Longrightarrow \quad |f''(x)| \leq 1 = K \end{array}$$

Thus,

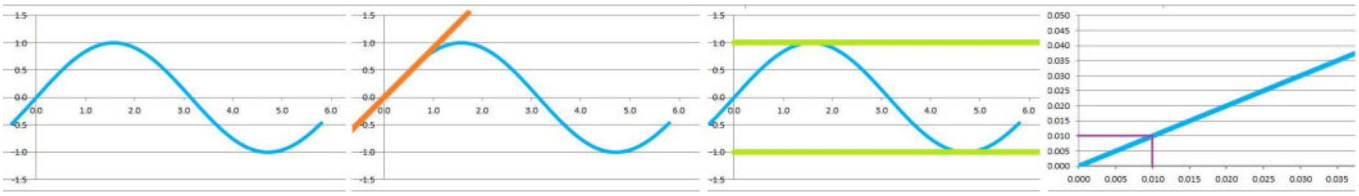
$$\sin .01 \approx .01\,,$$

and, furthermore, the accuracy is at worst

$$\varepsilon = \frac{1}{2}K(x-a)^2 = .5 \cdot 1 \cdot .01^2 = .00005\,.$$

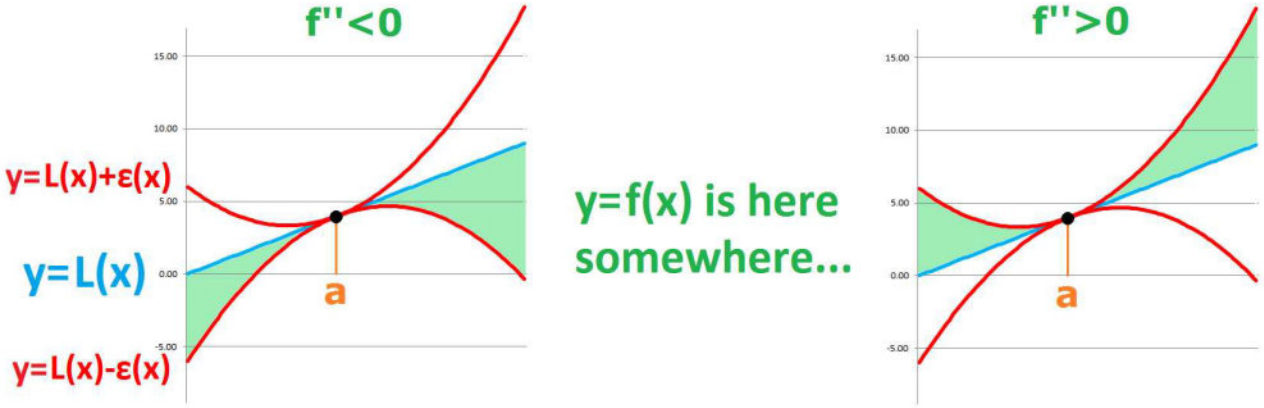
Therefore,

$$.01 - .00005 = .00995 \leq \sin .01 \leq .01005 = .01 + .00005\,.$$



Note that the choice of K in the last example was best possible. This was, therefore, a worst-case scenario. In contrast, here $K = 1$ isn't the best possible choice. By limiting our attention to the relevant part of the graph of $|f''|$ we discover a better value for the bound $K = .01$. This improves the accuracy by a factor of $.01$.

Knowing the concavity of the function that we approximate cuts the funnel in half.



Under the restrictions of the theorem, we conclude the following:

- When f is concave up (i.e., $f'' > 0$), we have for each x in $[a, b]$:

$$L(x) \leq f(x) \leq L(x) + \frac{1}{2}K(x-a)^2\,.$$

- When f is concave down (i.e., $f'' < 0$), we have for each x in $[a, b]$:

$$L(x) - \frac{1}{2}K(x-a)^2 \leq f(x) \leq L(x)\,.$$

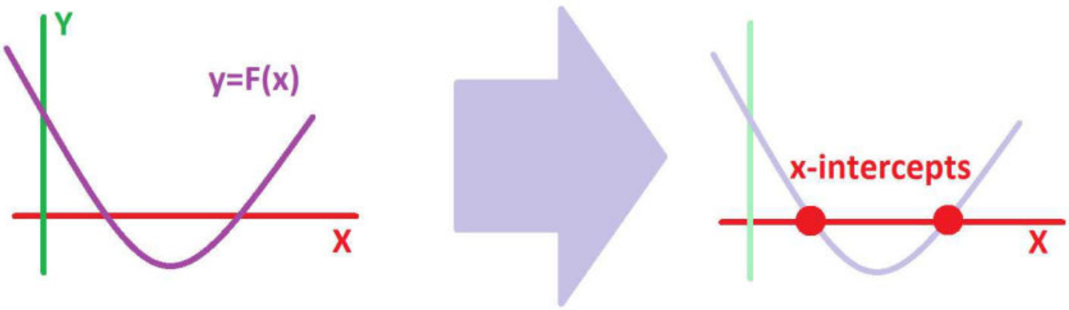
6.4. Solving equations numerically: bisection and Newton's method

What does it mean to *solve an equation*? Let's review (Volume 1, Chapter 1PC-5).

We start with:

▶ Given a function $y = f(x)$, find such a d that $f(d) = 0$.

Geometrically, we speak of the x-intercepts of the function:



Algebraic methods are precise but only apply to a very narrow class of functions (such as polynomials of degree below 5). What about *approximate solutions*? Solving the equation $y = f(x)$ "numerically" means the following:

▶ Find a sequence of numbers d_n such that $d_n \to d$ and $f(d) = 0$.

Then d_n are the *approximations* of the unknown d.

Let's recall the proof of the *Intermediate Value Theorem* (Chapter 1). We interpreted this proof as an iterated search for a solution of the equation $f(x) = 0$. We constructed a sequence of nested intervals by cutting them in half again and again. It is called *bisection*.

A function f is defined and is continuous on an interval $[a, b]$ with

$$f(a) < 0,\ f(b) > 0\,.$$

We want to approximate the following:

$$d \text{ in } [a, b] \text{ such that} f(d) = 0\,.$$

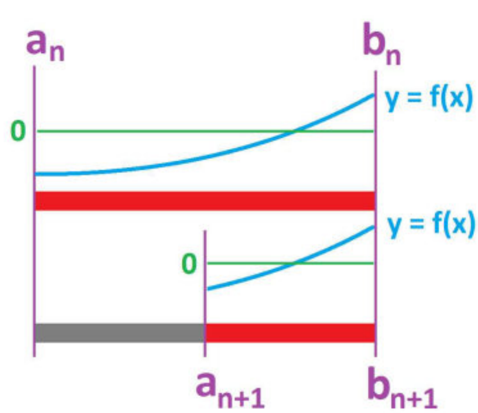
We start with the halves of $[a, b]$:

$$\left[a, \frac{1}{2}(a+b)\right] \quad \text{and} \quad \left[\frac{1}{2}(a+b), b\right]\,.$$

Over one of them (or both), the function f changes its sign. This fact can be expressed as follows:

$$f(a) \cdot f\left(\frac{1}{2}(a+b)\right) < 0 \quad \texttt{OR} \quad f\left(\frac{1}{2}(a+b)\right) \cdot f(b) < 0\,.$$

Whichever it is, we rename the ends of this interval a_1 and b_1:



Next, we consider the halves of this new interval and so on. We continue with this process and the result is two sequences of numbers:

$$a_n \quad \text{and} \quad b_n\,.$$

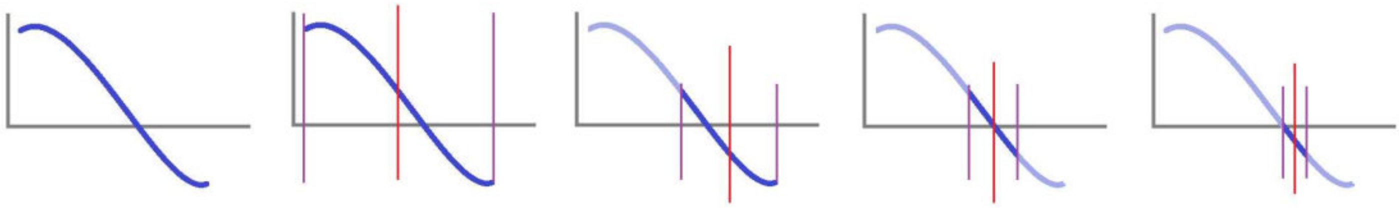
Exercise 6.4.1

Write a recursive formula for these intervals.

These two sequences of numbers form a "nested" sequence of intervals:

$$[a, b] \supset [a_1, b_1] \supset [a_2, b_2] \supset \dots$$

What is special about these intervals is that on each of them f changes its sign:



In other words, we have:

$$f(a_n) < 0,\ f(b_n) > 0 \quad \text{or} \quad f(a_n) > 0,\ f(b_n) < 0\,.$$

We conclude that the sequences converge to the same value,

$$a_n \to d,\ b_n \to d\,.$$

Furthermore, from the continuity of f, we conclude that the values of the function also converge:

$$f(a_n) \to f(d),\ f(b_n) \to f(d)\,.$$

Therefore, d is an (unknown) solution:

$$f(d) = 0\,.$$

The problem is solved!

Example 6.4.2: $\sin x = 0$

Let's review how the bisection method solves a specific equation:

$$\sin x = 0\,.$$

We started with the interval $[a_1, b_1] = [3, 3.5]$ and used the following spreadsheet formula for a_n:

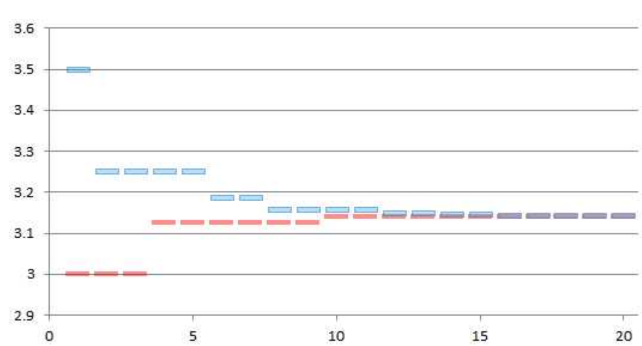
```
=IF(R[-1]C[3]*R[-1]C[4]<0,R[-1]C,R[-1]C[1])
```

and b_n:

```
=IF(R[-1]C[1]*R[-1]C[2]<0,R[-1]C[-1],R[-1]C)
```

The results are below:

n	An	Dn	Bn	f(An)	f(Dn)	f(Bn)	+/-
0	3	3.25	3.5	0.14112	-0.1082	-0.35078	-
1	3	3.25	3.25	0.14112	-0.1082	-0.1082	-
2	3	3.125	3.25	0.14112	0.016592	-0.1082	-
3	3.125	3.125	3.25	0.016592	0.016592	-0.1082	-
4	3.125	3.1875	3.25	0.016592	-0.04589	-0.1082	-
5	3.125	3.1875	3.1875	0.016592	-0.04589	-0.04589	-
6	3.125	3.15625	3.1875	0.016592	-0.01466	-0.04589	-
7	3.125	3.15625	3.15625	0.016592	-0.01466	-0.01466	-
8	3.125	3.140625	3.15625	0.016592	0.000968	-0.01466	-
9	3.140625	3.140625	3.15625	0.000968	0.000968	-0.01466	-
10	3.140625	3.148438	3.15625	0.000968	-0.00684	-0.01466	-
11	3.140625	3.148438	3.148438	0.000968	-0.00684	-0.00684	-
12	3.140625	3.144531	3.148438	0.000968	-0.00294	-0.00684	-
13	3.140625	3.144531	3.144531	0.000968	-0.00294	-0.00294	-
14	3.140625	3.142578	3.144531	0.000968	-0.00099	-0.00294	-
15	3.140625	3.142578	3.142578	0.000968	-0.00099	-0.00099	-
16	3.140625	3.141602	3.142578	0.000968	-8.9E-06	-0.00099	-
17	3.140625	3.141602	3.141602	0.000968	-8.9E-06	-8.9E-06	-
18	3.140625	3.141113	3.141602	0.000968	0.000479	-8.9E-06	-
19	3.141113	3.141113	3.141602	0.000479	0.000479	-8.9E-06	-
20	3.141113	3.141357	3.141602	0.000479	0.000235	-8.9E-06	-



The values of a_n, b_n are visibly approaching π (and so does d_n, the mid-point of the interval) while the values of $f(a_n)$, $f(b_n)$ approach 0.

Exercise 6.4.3

Solve the equation $\sin x = .2$.

There are other methods for solving $f(x) = 0$. One is to use the *linearization* of f as a substitute, one such approximation at a time.

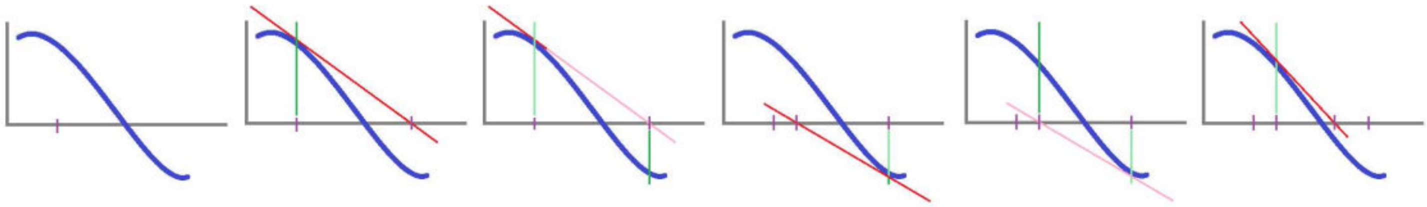
Suppose a function f is given as well as the initial estimate x_0 of a solution d of the equation $f(x) = 0$. We replace f in this equation with its linear approximation L at this initial point:

$$L(x) = f(x_0) + f'(x_0)(x - x_0)\,.$$

Then we solve the equation $L(x) = 0$ for x:

$$L(x) = f(x_0) + f'(x_0)(x - x_0) = 0\,.$$

In other words, we find the intersection of the tangent line with the x-axis:



The equation, which is *linear*, is easy to solve. The point of intersection is

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}\,.$$

Then we repeat the process for x_1. And so on...

This is called *Newton's method.* It is a sequence of numbers given recursively:

$$\boxed{x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}}$$

Warning!

The method fails when it reaches a point where the derivative is equal to (or even close to) 0.

Example 6.4.4: Newton's method

Let's use Newton's method to solve the same equation as above:

$$\sin x = 0\,.$$

The spreadsheet formula for x_n is as follows:

```
=R[-1]C-SIN(R[-1]C)/COS(R[-1]C)
```

We start with $x_0 = 3$. The sequence converges to π very quickly (left):

n	Xn
0	3
1	3.1425465430743
2	3.1415926533005
3	3.1415926535898
4	3.1415926535898
5	3.1415926535898
6	3.1415926535898
7	3.1415926535898
8	3.1415926535898
9	3.1415926535898
10	3.1415926535898

n	Xn
0	2
1	4.1850398632615
2	2.4678936745147
3	3.2661862775691
4	3.1409439123176
5	3.1415926536808
6	3.1415926535898
7	3.1415926535898
8	3.1415926535898
9	3.1415926535898
10	3.1415926535898

n	Xn
0	1.7
1	9.3966021394592
2	9.4247854191824
3	9.4247779607694
4	9.4247779607694
5	9.4247779607694
6	9.4247779607694
7	9.4247779607694
8	9.4247779607694
9	9.4247779607694
10	9.4247779607694

n	Xn
0	pi()/2
1	#VALUE!
2	#VALUE!
3	#VALUE!
4	#VALUE!
5	#VALUE!
6	#VALUE!
7	#VALUE!
8	#VALUE!
9	#VALUE!
10	#VALUE!

However, as our choice of x_0 changes and gets closer to $\pi/2$, the value of x_1 becomes larger and larger. In fact, it might be so large that the sequence won't ultimately converge to π but to 2π, 100π, etc. If we choose $x_0 = \pi/2$ exactly, the tangent line is horizontal and the algorithm breaks down!

Exercise 6.4.5

Solve the equation $\sin x = .2$.

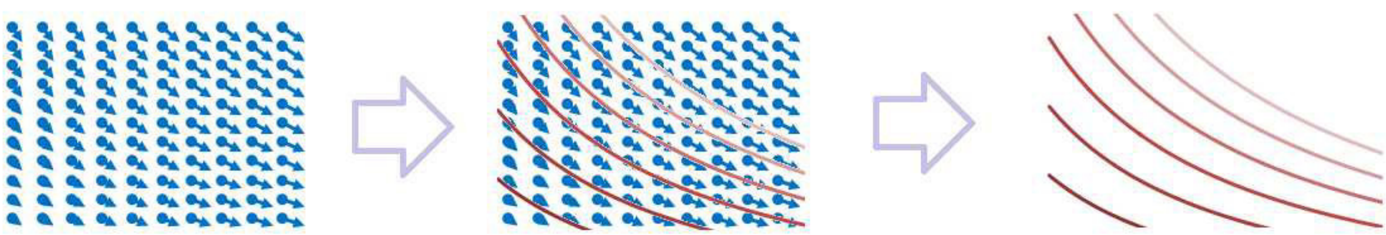
6.5. *Particle in a flow*

Previously, we considered an example of how functions appear as direct representations of the dynamics given by an indirect description.

For example, an object's velocity is derived from the acceleration, and then the location from the velocity. The acceleration (or the force) is a function of *time*, and so is the velocity.

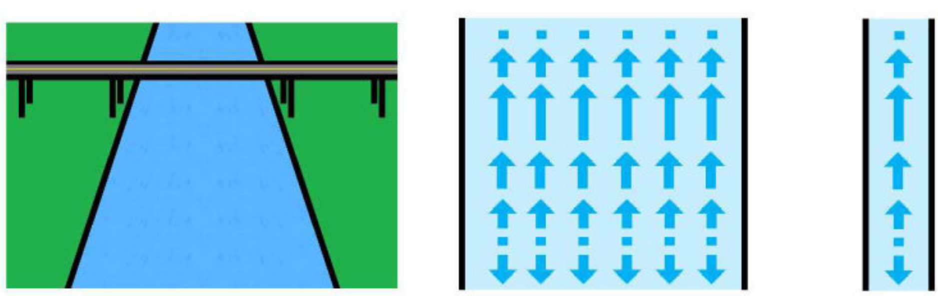
A different, and just as important, example of such emergence is flows of liquids. The difference is that here the velocity is a function of *location*.

The task is to determine the location of a particle (as a function of time) in a flow of liquid:



We start with a discrete model and limit ourselves to the 1-dimensional case: *pipes and canals*.

Suppose there is a pipe with the velocity of the stream measured somehow at each location. Similarly, this is a canal with the water that has the exact same velocity – parallel to the canal – at all locations across it. In other words, the velocity only varies along the length of the pipe or the canal:



That's what makes the problem one-dimensional.

Our goal is the following:

▶ Trace a single particle of this stream.

We will simply apply the familiar formula to compute the location from the velocity:

$$\text{displacement} \ = \ \text{velocity} \ \cdot \ \text{time increment}\,.$$

A fixed time increment Δt is supplied ahead of time even though in general it can also be variable.

We start with the following two quantities provided by the model we are to implement:

- the initial time t_0, and
- the initial location p_0.

They are placed in the first row of the table; for example:

	iteration n	time t_n	velocity v_n	location p_n
initial:	0	3.5	$--$	22

This is the starting point. We would like to know the values of these quantities at every moment of time, in these increments.

As we progress in time and space, new numbers are placed in the next row of our spreadsheet. This is how the second row, $n = 1,\ t_1 = t_0 + \Delta t$, is finished.

The current velocity v_1 is in the first cell of the row and the initial location p_0 is in the last cell of the last row. The following *recursive formula* is placed in the second cell of the new row of our spreadsheet:

▶ next location = initial location + current velocity · time increment.

We continue with the rest in the same manner.

Below is the recursive algorithm for computing row $n + 1$ from row n:

First column is the time t_n: Next moment of time = last moment of time + time increment. $t_{n+1} = t_n + \Delta t$
Second column is the velocity v_n: Next velocity explicitly depends on the values in the previous row. $v_{n+1} =$ any combination of $t_n,\ v_n,\ p_n$
Third column is the location p_n: Next location = last location + current velocity · time increment. $p_{n+1} = p_n + v_{n+1} \cdot \Delta t$

This dependence is shown below for $\Delta t = .1$:

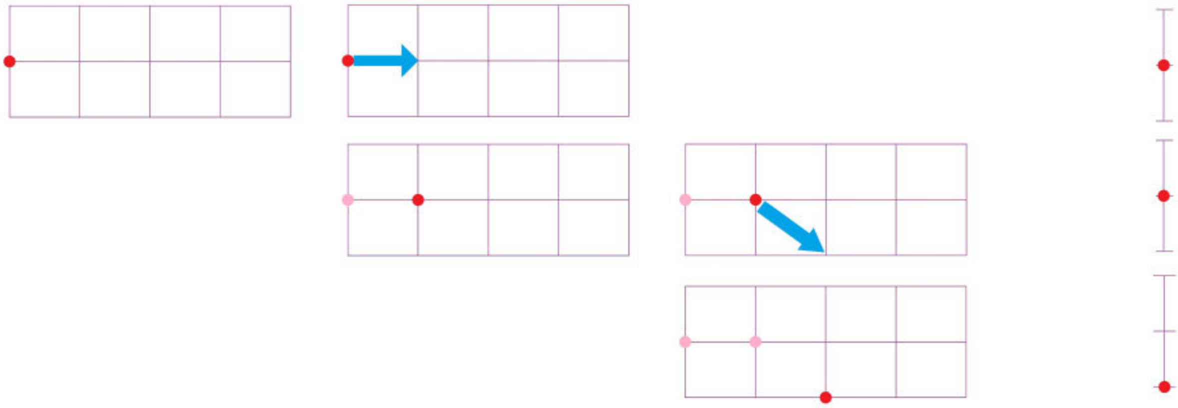
$$\begin{array}{r|clclcl|}
 & \text{iteration } n & \text{time } t_n & & \text{velocity } v_n & & \text{location } p_n \\ \hline
\text{initial:} & 0 & t_0 = 3.5 & & -- & & p_0 = 22 \\
 & & \downarrow & & & \swarrow & \downarrow \\
 & 1 & t_1 = 3.5 + .1 & \to & v_1 = 33 & \to & p_1 = 22 + 33 \cdot .1 \\
 & & \downarrow & & & \swarrow & \downarrow \\
 & 2 & t_2 = ? & \to & v_2 = ? & \to & p_2 = ? \\
 & & \downarrow & & & \swarrow & \downarrow
\end{array}$$

Also, in a flow, the current velocity of a particle depends – somehow – on the current time or its (last) location, as indicated by the arrows. This dependence may be an explicit formula or it may come from the instruments' readings.

As we progress in time and space, a number is supplied and are placed in each of the columns of our spreadsheet one row at a time:

$$t_n,\ v_n,\ p_n,\ n = 1, 2, 3, \ldots$$

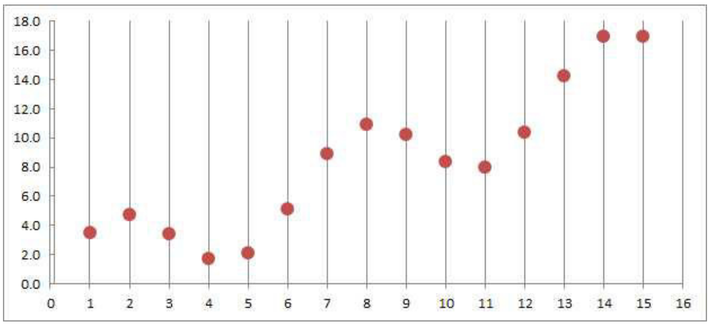
We then plot the time and location on the plane with axes t (horizontal) and p (vertical). The result is a sequence of points developing one row at a time that might look like this:



The result is a growing table of values:

	iteration n	time t_n	velocity v_n	location p_n
initial:	0	3.5	$--$	22
	1	3.6	33	25.3
	...	...	...	...
	1000	103.5	4	336
	...	...	...	...

The result may be seen as three sequences $t_n,\ v_n,\ p_n$ or as the table of values of two *functions* of t. The graph of the position might look like this:



Let's test a few instances of this algorithm with a spreadsheet. We use the following formula for the position:

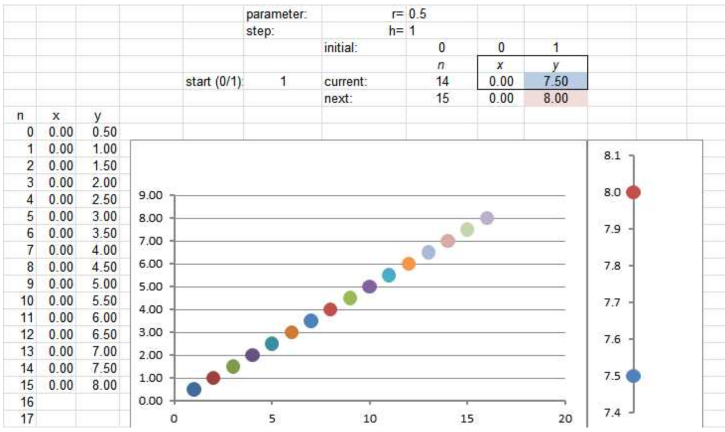
```
=R[-1]C+RC[-1]*(RC[-2]-R[-1]C[-2])
```

Example 6.5.1: constant velocity

Suppose that the velocity of the stream in the pipe is constant:

$$v_{n+1} = .5\,.$$

The simulation shows that the particle progresses in a uniform fashion:



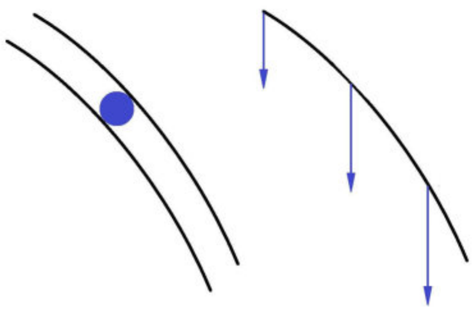
n	x	y
0	0.00	0.50
1	0.00	1.00
2	0.00	1.50
3	0.00	2.00
4	0.00	2.50
5	0.00	3.00
6	0.00	3.50
7	0.00	4.00
8	0.00	4.50
9	0.00	5.00
10	0.00	5.50
11	0.00	6.00
12	0.00	6.50
13	0.00	7.00
14	0.00	7.50
15	0.00	8.00
16		
17		

Exercise 6.5.2

Find the recursive formula for the position.

Example 6.5.3: curved pipe

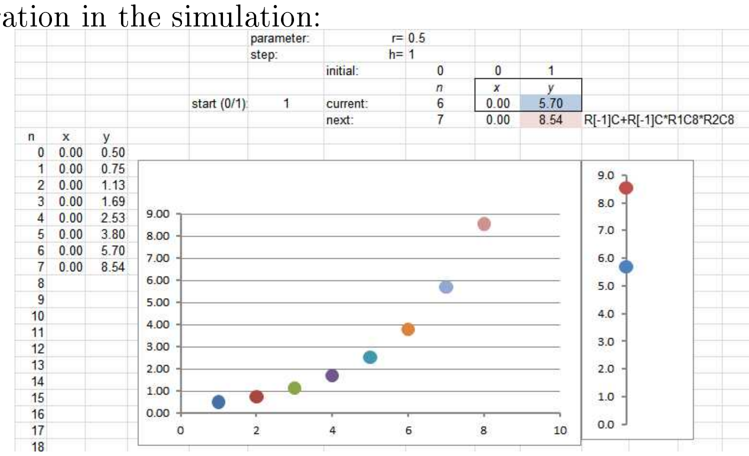
Suppose that there is a source of water in the middle of the pipe but it is also bent down so that the speed is higher away from this location:



Specifically, we assume that the velocity of the stream is directly proportional to the distance from the source:

$$v_{n+1} = .5 \cdot p_n \,.$$

We can see the acceleration in the simulation:



n	x	y
0	0.00	0.50
1	0.00	0.75
2	0.00	1.13
3	0.00	1.69
4	0.00	2.53
5	0.00	3.80
6	0.00	5.70
7	0.00	8.54
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		

Exercise 6.5.4

Find the recursive formula for the position.

Exercise 6.5.5

Modify the formulas for the case of a variable time increment: $\Delta t_{n+1} = t_{n+1} - t_n$.

Next, a *continuous flow*.

What is the difference? The time t varies over a whole interval $[t_0, \infty)$ and the particle is progressing

continuously through an interval of space $p = y(t)$. This is the answer; however, what is the question?

We pose the following problem:

> ▶ Suppose the velocity is given by an explicit formula as a function of the location $z = f(y)$ defined on an interval J of space. Is there an explicit formula for the location as a function of time $y = y(t)$ defined on an interval I of time?

We start from the discrete case.

First, we recast our recursive relation,

$$p_{n+1} = p_n + v_{n+1} \cdot \Delta t\,,$$

as if this sequence of locations comes from *sampling* our function y:

$$p_n = y(t_n)\,.$$

Then, we also have:

$$v_n = f(p_n)\,.$$

This holds for *every* $\Delta t > 0$. Here f is known while y is unknown.

Second, with $t = t_n$, the recursive formula becomes:

$$y(t + \Delta t) = y(t) + f(y(t + \Delta t)) \cdot \Delta t\,.$$

Therefore,

$$\frac{y(t + \Delta t) - y(t)}{\Delta t} = f(y(t + \Delta t))\,.$$

Taking the limit over $\Delta t \to 0$ gives us the following:

$$y'(t) = f(y(t))\,,$$

provided the limits exist of course.

The above equation is called a *differential equation*. Its abbreviated version is below:

$$\boxed{y' = f(y)}$$

The attitude is identical to that for anti-differentiation. We again think of an *equation*, an equation for functions:

Solve for y	Solve for y
$\dfrac{dy}{dx} = x^2$	$\dfrac{dy}{dx} = y^2$

The latter is obviously more complex!

Such equations are considered below and throughout the rest of calculus.

The typical requirements for the functions involved are the following:

- $y = y(t)$ is differentiable on the interval I.
- $z = f(y)$ is continuous on the interval J.

Example 6.5.6: curved pipe, continued

If the pipe's slope varies from location to location, the velocity of the flow will depend on the location too. For example, if the pipe is curved down, the velocity of the flow will be higher the farther away the point is from the origin. Let's again consider the case when the velocity is proportional to the location:

$$v_{n+1} = .5 \cdot p_n \,.$$

For the continuous case, we have a differential equation:

$$y' = .5 \cdot y \,.$$

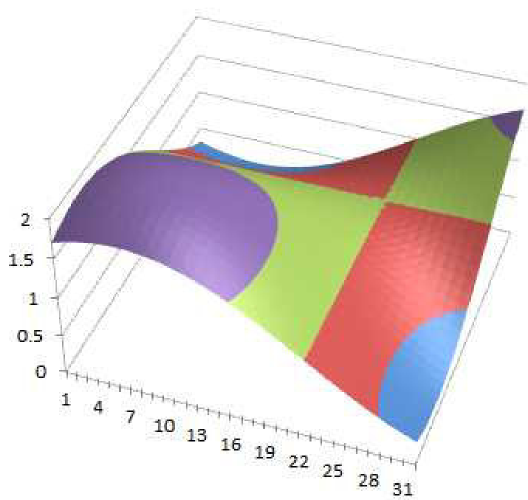
We already know its solution:

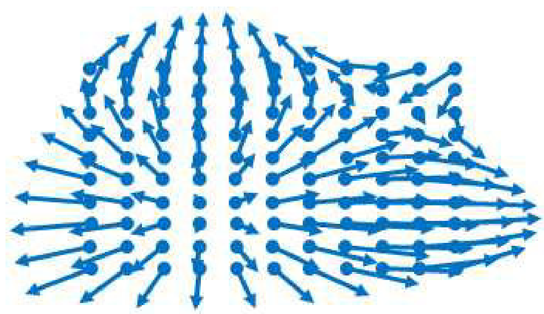
$$y(t) = Ce^{.5t} \,,$$

for any C.

Let's consider the discrete case of *flows on the plane* now.

One can imagine the flow of rainwater down a mountain terrain:



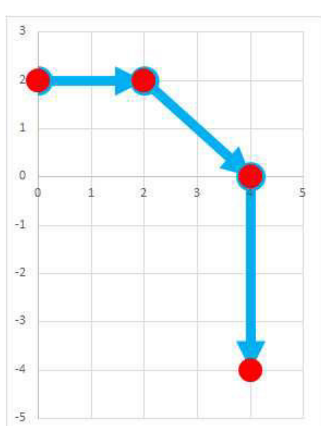


One can also imagine that this is a ball rolling down the slope following these directions:

Computationally, instead of two (velocity – location), our table will have four main columns:

iteration	time	velocity	velocity	location	location	...
n	t	v	x	u	y	...
0	3.5	$--$	22	$--$	3	...
1	3.6	33	25.3	4	3.5	...
...	...	...	...	...	...	...

To sketch the path of a particle in the stream, we only show the two spatial coordinates and hide the time:



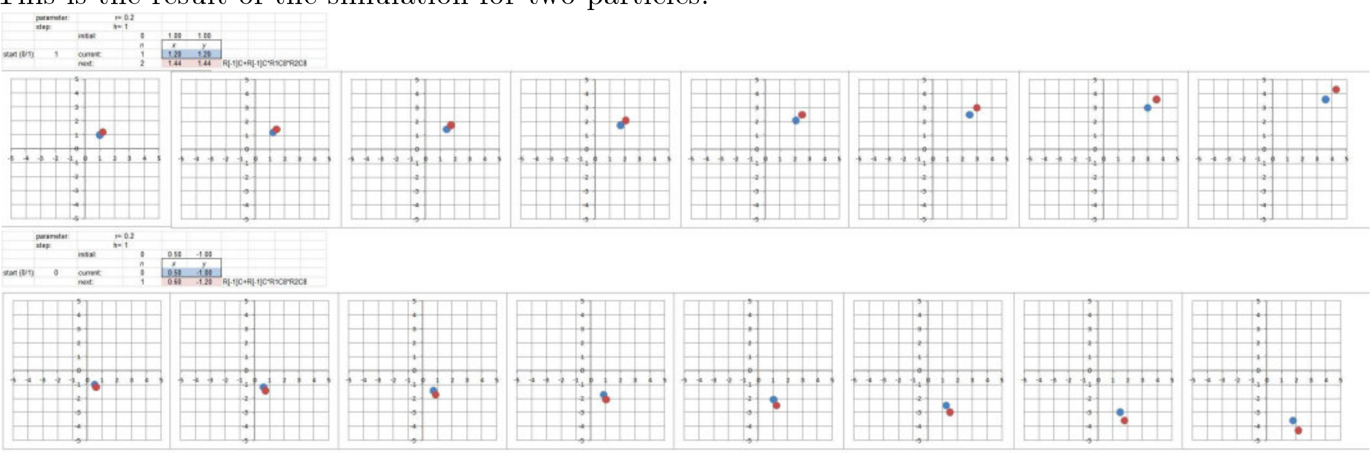
Example 6.5.7: curved surface

Just as with the pipe, if the terrain's slope varies from location to location, the velocity of the flow will depend on the location too. For example, if the terrain is curved down, the velocity of the flow will be higher the farther away the point is from the top of the mountain:

Let the velocity be proportional to the location for both horizontal and vertical.:

$$v_{n+1} = .2 \cdot x_n \quad \text{and} \quad u_{n+1} = .2 \cdot y_n \,.$$

This is the result of the simulation for two particles:



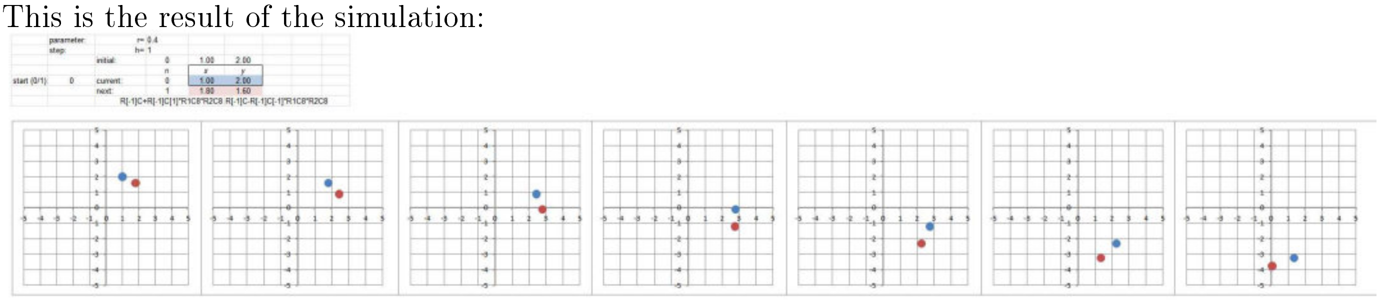
The particles are running in the direction away from the center, faster and faster.

Example 6.5.8: whirl

For more complex patterns, the vertical and horizontal will be interdependent. For example, the horizontal velocity may be proportional to the vertical location and the vertical velocity proportional to the *negative* of the horizontal location:

$$v_{n+1} = .2 \cdot y_n \quad \text{and} \quad u_{n+1} = -.2 \cdot x_n \,.$$

This is the result of the simulation:



Exercise 6.5.9

What are the continuous models for the flows described in the examples? Solve them.

Exercise 6.5.10

Suggest models of flow for other locations on this surface.

6.6. Differential equations

In this section, we will review and summarize our encounters with differential equations.

A *differential equation* is an equation that relates the values of the derivative of the function to the function's values.

Example 6.6.1: zero derivative

The simplest example of a differential equation is:

$$f'(x) = 0 \text{ for all } x\,.$$

We already know *the* solution:

- Constant functions are solutions.
- Conversely, only constant functions are solutions.

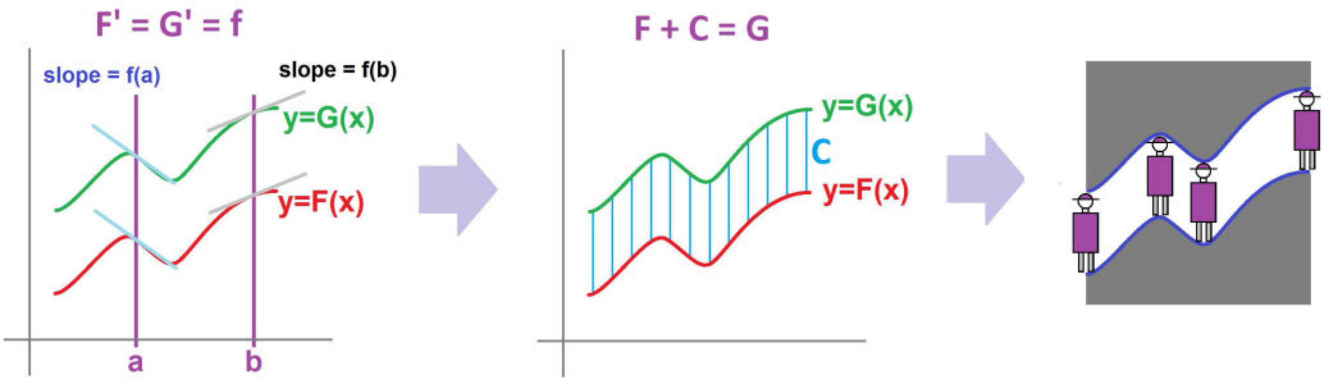
Example 6.6.2: anti-differentiation

We can replace 0 with any function:

$$f'(x) = g(x) \ \text{ for all } x\,.$$

We already know the solution:

- A solution f has to be an antiderivative of g.
- Conversely, only antiderivatives of g are solutions.



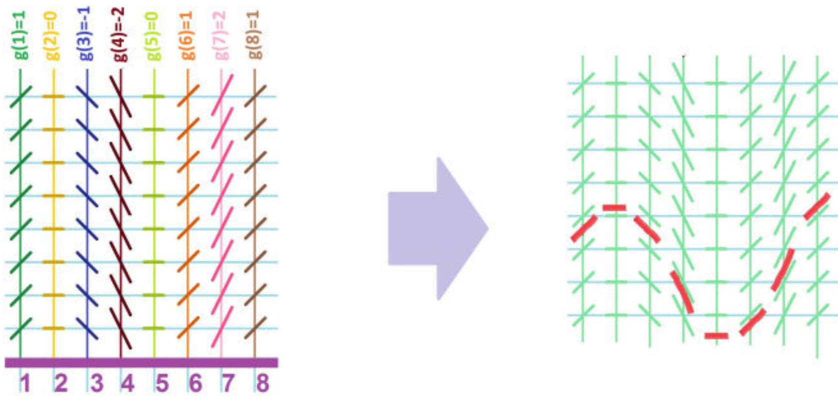
In summary, a differential equation is an equality of functions that may include derivatives. To solve the equation is to find all possible functions that satisfy it.

Let's take a look at the problem with a fresh eye.

For the first kind of differential equation, we will be looking for functions $y = y(x)$ that satisfy the equation,

$$y'(x) = g(x) \ \text{ for all } x\,.$$

What we know from the equation is the value of the derivative $y'(x)$ of y at every point x, but we don't know the value $y(x)$ of the function itself. Then, for every x, we know the slope of the tangent line at $(x, y(x))$. As $y(x)$ is unknown, in order to visualize the data, we plot the same slope for every point (x, y) on a given *vertical* line:



Thus, for each $x = c$, we indicate the angle α, with $g(c) = \tan\alpha$, of the intersection of the graph of the unknown function $y = y(x)$ and the vertical line $x = c$.

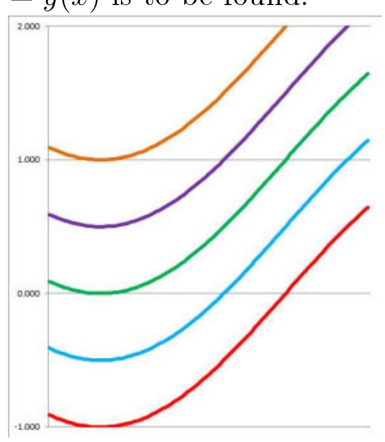
It is as if we are to create a *fabric* from two types of threads. The *vertical* ones have already been placed, and the way at which the threads of the second type are to be weaved in has been indicated.

The challenge is to devise a function that would *cross these lines at these exact angles.*

Is it always possible to have these functions? Yes, at least when g is continuous. How do we find them? Anti-differentiation.

Example 6.6.3: position from velocity

A familiar interpretation is that of an object the velocity $v(x) = y'(x)$ of which is known at any moment of *time* x and its location $y = y(x)$ is to be found.



These solutions *fill* the plane without intersections. They are just the vertically shifted versions of one of them.

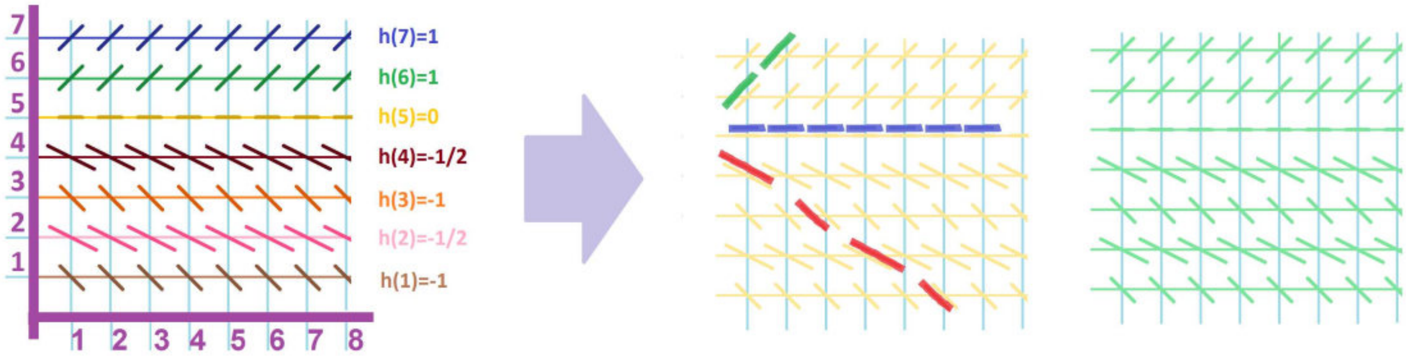
Another kind of a differential equation is more challenging:

- ► The derivative depends on the values of the function.

We are looking for functions $y = y(x)$ that satisfy the equation,

$$y'(x) = h(y(x)) \quad \text{for all } x\,.$$

What we know from the equation is the value of the derivative $y'(x)$ of y at every point (x, y) even though we don't know the value $y(x)$ of the function itself. Then, for every y, we know the slope of the tangent line at (x, y). As $y(x)$ is unknown, in order to visualize the data, we plot the same slope for every point (x, y) on a given *horizontal* line:



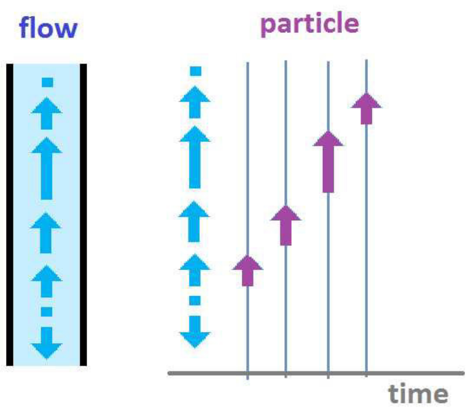
Thus, for each $y = d$, we indicate the angle α, with $h(d) = \tan\alpha$, of the intersection of the graph of the unknown function $y = y(x)$ and the vertical line $x = c$ with $y(c) = d$.

It is as if we are to create a fabric from two types of threads. The *horizontal* ones have already been placed and the way at which the threads of the second type are to be weaved in has been indicated.

Is it always possible to have these functions? Yes, at least when h is differentiable. The challenge is to devise a function that would *cross these lines at these exact angles*. How do we find them? It is hard or impossible (Volume 5, Chapter 5DE-1).

Example 6.6.4: flow

An interpretation is a stream of liquid with its velocity known at every *location* y and we need to trace the path of a particle initially located at a specific place y_0.



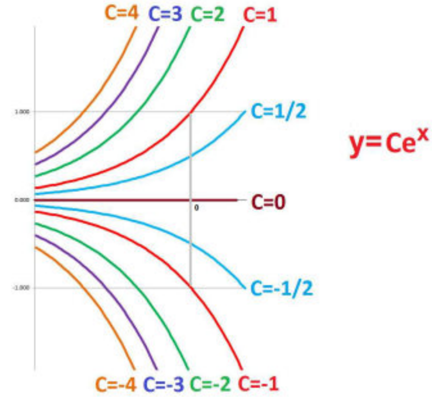
Example 6.6.5: population growth

This is what happens when the velocity is equal to the location:

$$y' = y\,.$$

The population doubles every unit of time!

To solve, what is the function equal to its derivative? It's the exponent $y = e^x$, of course. However, all of its multiples $y = Ce^x$ are also solutions:



Once again, these solutions *fill* the plane without intersecting.

Exercise 6.6.6

What is the transformation of the plane that creates all these curves from one?

Exercise 6.6.7

Can the velocity be really "equal" to the location?

Exercise 6.6.8

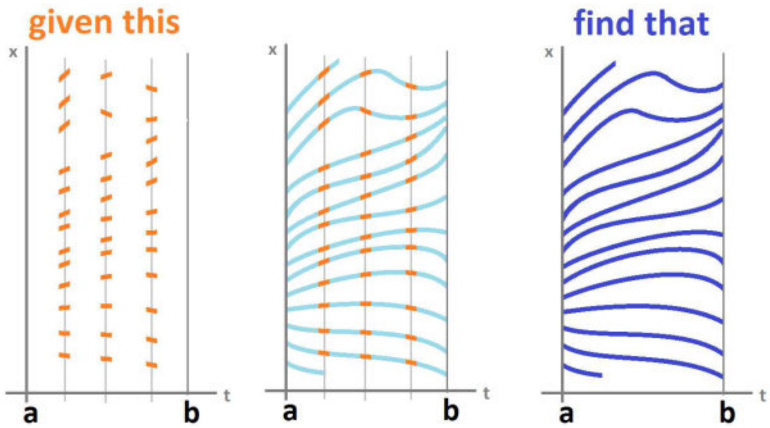
What happens when the velocity is proportional to the location; i.e.,

$$y' = ky?$$

Compare and contrast:

DE:	$y'(x) = g(x)$	$y'(x) = h(y(x))$
the slopes are the same along any	vertical line	horizontal line
the velocity is known at any	time	location

Of course, also common are differential equations that have neither of these patterns:



Even though we don't consider the methods of solving these equations here (see Volume 5, Chapter 5DE-1), once we have a candidate function, it is easy to test it.

Example 6.6.9: antidifferentiation

This has been our approach to antidifferentiation so far. Solve:

$$y'(x) = xe^{x^2}\,.$$

Let's try:

$$y = \frac{1}{2}e^{x^2}\,.$$

Differentiation (with the Chain Rule) confirms that this is, indeed, a solution:

$$y'(x) = \left(\frac{1}{2}e^{x^2}\right)' = \frac{1}{2}2xe^{x^2} = xe^{x^2}\,.$$

Example 6.6.10: migration

Consider:

$$y' = my + b,\ m \neq 0\,.$$

It can represent population growth/decay accompanied by *migration* at a constant rate. Try:

$$y(x) = -\frac{b}{m} + Ce^{mt}\,,$$

where C is any real number. We substitute this expression into the left-hand side of the equation:

$$\left(-\frac{b}{m} + Ce^{mt}\right)' = \left(Ce^{mt}\right)' = Cme^{mt}\,.$$

We now substitute it into the right-hand side of the equation:

$$m\left(-\frac{b}{m} + Ce^{mt}\right) + b = -b + mCe^{mt} + b = Cme^{mt}\,.$$

Confirmed!

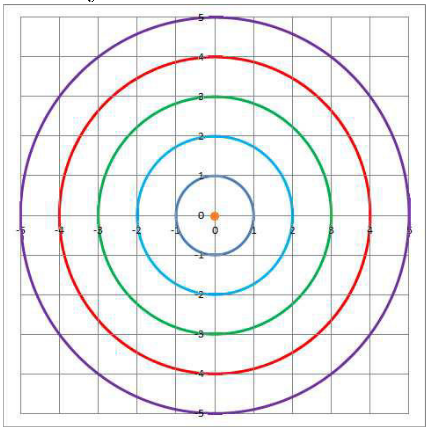
Exercise 6.6.11

Confirm that the function $y = r + Ce^{-kt}$ is a solution of the differential equation $y' = k \cdot (r - y)\,.$

Furthermore, once we have a family of curves, it is easy to find the differential equation it came from.

Example 6.6.12: circles

What differential equation does the family of all concentric circles around 0 satisfy?



This family is given by:

$$x^2 + y^2 = r^2, \quad \text{real } r\,.$$

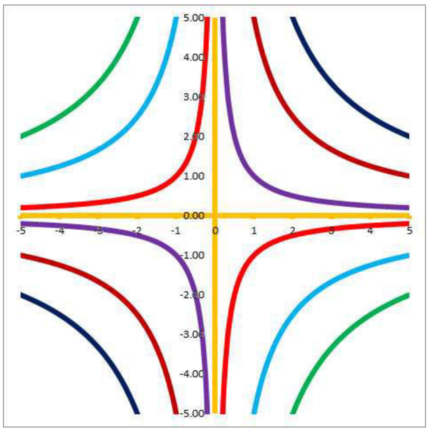
We simply differentiate the equation implicitly (Chapter 4):

$$x^2 + y^2 = r^2 \implies 2x + 2yy' = 0\,.$$

That's the equation.

Example 6.6.13: hyperbolas

What about this family of hyperbolas?



This family is given by:

$$xy = C, \quad \text{real } C\,.$$

Again, we differentiate the equation implicitly. We have:

$$y + xy' = 0\,.$$

Exercise 6.6.14

What is the differential equation for these parabolas $y = x^2 + C$? What about $y = Cx^2$? Or $x + C = y^2$?

6.7. *Motion under forces*

In this section we will summarize that we have learned about modeling motion.

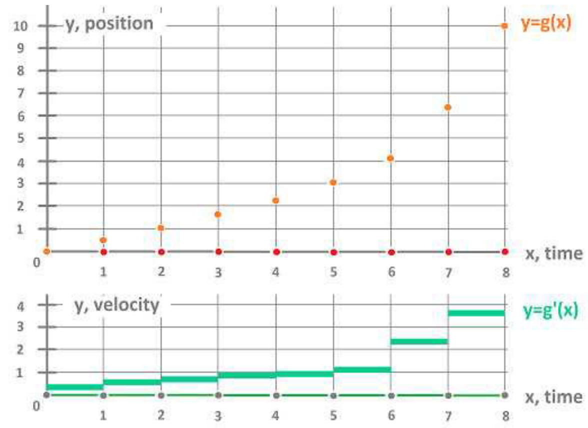
The list of the main quantities of the Newtonian physics, in dimension 1, is presented below:

Definition 6.7.1: quantities of Newtonian physics

Suppose a function r is defined at the primary nodes of a partition of a closed interval or on an open interval. If the independent variable t is called "time" and the function r is called "location" or "position", then the related quantities are defined as functions of time and denoted as follows:

quantity	incremental	nodes	continuous	physics notation
location:	r	primary	r	
displacement:	$D = \Delta r$	secondary	$D = dr$	
velocity:	$v = \dfrac{\Delta r}{\Delta t}$	secondary	$v = \dfrac{dr}{dt}$	$= \dot{r}$
momentum:	$p = mv$	secondary	$p = mv$	
impulse:	$J = \Delta p$	primary	$J = dp$	
acceleration:	$a = \dfrac{\Delta^2 r}{\Delta t^2}$	primary	$a = \dfrac{d^2 r}{dt^2}$	$= \ddot{r}$

For example, the upward concavity of the function below is obvious, which indicates a positive acceleration:



Let's state some elementary facts about the Newtonian physics.

Newton's First Law

If the net force is zero, then the velocity v of the object is constant:

$$\boxed{F = 0 \implies v = \text{ constant}}$$

The law can be restated without invoking the geometry of time. If the net force is zero, then the displacement Δr of the object is constant:

$$F = 0 \implies \Delta r = \text{ constant}\ .$$

The law shows that the only possible type of motion in such a force-less environment is uniform; i.e., it is a repeated addition:

$$r(t + \Delta t) = r(t) + c\,.$$

Newton's Second Law

The net force on an object is equal to the derivative of its momentum p:

$$\boxed{F = \frac{\Delta p}{\Delta t} \quad \text{or} \quad F = \dot{p}}$$

Newton's Third Law

If one object exerts a force F_1 on another object, the latter simultaneously exerts a force F_2 on the former, and the two forces are exactly opposite:

$$\boxed{F_1 = -F_2}$$

Theorem 6.7.2: Law of Conservation of Momentum

Suppose a system of objects is "closed":

- *There is no exchange of matter with its surroundings.*
- *There are no external forces.*

Then the total momentum is constant:

$$\boxed{p = \text{ constant}}$$

In other words,

$$J = dp = 0\,.$$

Proof.

Consider two objects interacting with each other. By the Third Law, the forces between them are exactly opposite:

$$F_1 = -F_2 \,.$$

Due to the Second Law, we conclude that

$$p_1' = -p_2' \,.$$

From the Sum Rule, we have:

$$(p_1 + p_2)' = 0 \,.$$

Exercise 6.7.3

State the equation of motion for a variable-mass system (such as a rocket). Hint: Apply the second law to the entire, constant-mass system.

Exercise 6.7.4

Create a spreadsheet that computes all of these quantities.

Suppose we know the forces affecting a moving object. How can we predict its dynamics?

Assuming a fixed mass, the total force gives us our acceleration. Then, we apply the same formula (i.e., anti-differentiation) to compute:

- the velocity from the acceleration, and then
- the location from the velocity.

We already know how to find the location from the velocity according to the analysis of motion of a fluid flow earlier in this chapter. This will be a follow-up.

Below we examine the *discrete model* of motion.

A fixed time increment Δt is supplied ahead of time even though it can also be variable.

We start with the following three quantities that come from the setup of the motion:

- the initial time t_0,
- the initial velocity v_0, and
- the initial location p_0.

They are placed in the four consecutive cells of the first row of the table:

	iteration n	time t_n	acceleration a_n	velocity v_n	location p_n
initial:	0	3.5	$--$	33	22

Another quantity that comes from the setup is

▶ the current acceleration a_1.

It will be placed in the next row.

This is the starting point. We would like to know the values of all of these quantities at every moment of time, in these increments.

As we progress in time and space, new numbers are placed in the next row of our table. This is how the second row, $n = 1,\ t_1 = t_0 + \Delta t$, is completed.

The current acceleration a_0 given in the first cell of the second row. The current velocity v_1 is found and placed in the second cell of the second row of our table:

▶ current velocity = initial velocity + current acceleration · time increment.

The second quantity we use is the initial location p_0. The following is placed in the third cell of the second row:

▶ current location = initial location + current velocity · time increment.

This dependence is shown below:

	iteration n	time t_n	acceleration a_n		velocity v_n		location p_n
initial:	0	3.6	$--$		33		22
					$\downarrow$		$\downarrow$
current:	1	t_1	66	$\to$	v_1	$\to$	p_1

These are *recursive formulas*, just as before.

We continue with the rest in the same manner. As we progress in time and space, numbers are supplied and placed in each of the four columns of our spreadsheet one row at a time:

$$t_n,\ a_n,\ v_n,\ p_n,\ n = 1, 2, 3, \ldots$$

The first quantity in each row we compute is the time:

$$t_{n+1} = t_n + \Delta t\,.$$

The next is the acceleration a_{n+1} which may be constant (such as in the case of a free-falling object) or may explicitly depend on the values in the previous row.

Where does the current acceleration come from? It may come as pure *data*: The column is filled with numbers ahead of time or it is being filled as we progress in time and space.

There may also be an explicit, functional dependence of the acceleration (or the force) on the rest of the quantities. The acceleration may depend on the following:

1. the current time, e.g., $a_{n+1} = \sin t_{n+1}$ such as when we speed up the car, or
2. the last location, e.g., $a_{n+1} = 1/p_n^2$ such as when the gravity depends on the distance to the planet, or
3. the last velocity, e.g., $a_{n+1} = -v_n$ such as when the air resistance works in the opposite direction of the velocity,
4. or all three.

Exercise 6.7.5

Draw arrows in the above table to illustrate these dependencies.

Simple examples of case 1 above are addressed below. More examples of case 1 are discussed in Volume 3, Chapter 3IC-1, and in the multidimensional setting in Volume 4, Chapter 4HD-2. Case 2 and case 3 are considered further in Volume 5, Chapter 5DE-1.

The nth iteration of the velocity v_n is computed:

- Current velocity = last velocity + current acceleration · time increment,
- $v_{n+1} = v_n + a_n \cdot \Delta t$.

The values of the velocity are placed in the second column of our table.

The nth iteration of the location p_n is computed:

- Current location = last location + current velocity · time increment,
- $p_{n+1} = p_n + v_n \cdot \Delta t$.

The values of the location are placed in the third column of our table.

The result is a growing table of values:

	iteration n	time t_n	acceleration a_n	velocity v_n	location p_n
initial:	0	3.5	$--$	33	22
	1	3.6	66	38.5	25.3
	...	...	...	...	...
	1000	103.5	666	4	336
	...	...	...	...	...

The result may be seen as four sequences $t_n,\ a_n,\ v_n,\ p_n$ or as the table of values of three *functions* of t.

Exercise 6.7.6

Implement a variable time increment: $\Delta t_{n+1} = t_{n+1} - t_n$.

Example 6.7.7: rolling ball

A rolling ball is unaffected by horizontal forces. Therefore, $a_n = 0$ for all n. The recursive formulas for the horizontal motion simplify as follows:

- The velocity $v_{n+1} = v_n + a_n \cdot \Delta t = v_n = v_0$ is constant.
- The position $p_{n+1} = p_n + v_n \cdot \Delta t = p_n + v_0 \cdot \Delta t$ grows at equal increments.

In other words, the position depends linearly on the time.

Example 6.7.8: falling ball

A falling ball is unaffected by horizontal forces and the vertical force is constant: $a_n = a$ for all n. The first of the two recursive formulas for the vertical motion simplifies as follows:

- The velocity $v_{n+1} = v_n + a_n \cdot \Delta t = v_n + a \cdot \Delta t$ grows at equal increments.
- The position $p_{n+1} = p_n + v_n \cdot \Delta t$ grows at linearly increasing increments.

It follows that the position depends quadratically on the time.

We now turn to motion in *dimension* 2 such as in the case of angled flight.

One considers two functions for each of the above quantities: vertical and horizontal components. Furthermore, instead of three (acceleration – velocity – location), there will be six main columns :

	time	horiz. acceleration	horiz. velocity	horiz. position	vert. acceleration	vert. velocity	vert. position	...
n	t	a_n	v_n	x_n	b_n	u_n	y_n	...
0	3.5	$--$	33	22	-10	5	3	...
1	3.6	66	38.5	25.3	-15	4	3.5	...
...	...	...	...	...	...	...	...	...
1000	103.5	666	4	336	14	66	4	...
...	...	...	...	...	...	...	...	...

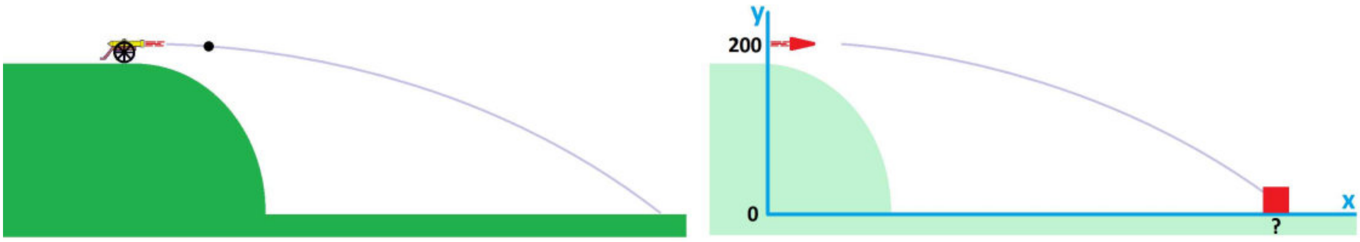
We would have nine columns when the model is three-dimensional, such as in the case of an object thrown with a side wind.

Example 6.7.9: cannon

A falling ball is unaffected by horizontal forces and the vertical force is constant:

$$x: \ a_{n+1} = 0; \quad y: b_{n+1} = -g\,.$$

Now recall the setup considered previously: From a 200 feet elevation, a cannon is fired horizontally at 200 feet per second.



The initial conditions are for x and y respectively:

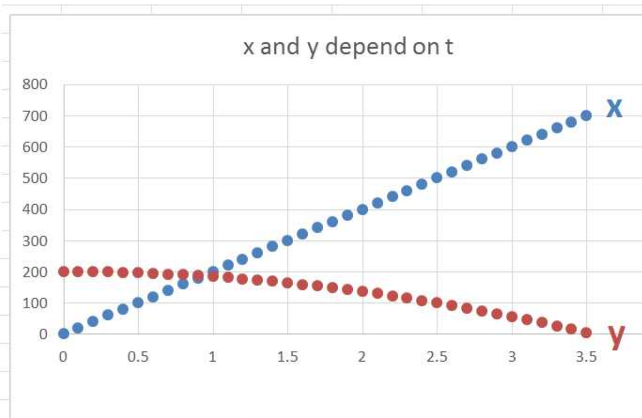
- The initial location is given by: $x_0 = 0$ and $y_0 = 200$.
- The initial velocity is given by: $v_0 = 200$ and $yu_0 = 0$.

Then we have two pairs of recursive equations independent of each other:

$$\begin{array}{r|lllllll} x: & v_{n+1} & = v_0, & & x_{n+1} & = x_n & +v_n\Delta t \\ y: & u_{n+1} & = u_n & -g\Delta t & y_{n+1} & = y_n & +u_n\Delta t \end{array}$$

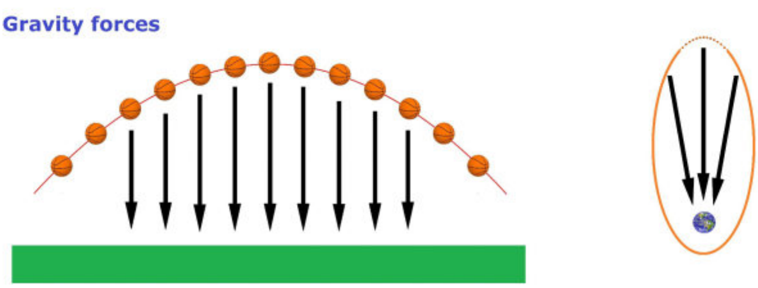
Implemented with a spreadsheet, the formulas produce the same results as the explicit formulas did before:

time	depth x	height y
0	0	200
0.1	20	199.84
0.2	40	199.36
0.3	60	198.56
0.4	80	197.44
0.5	100	196
0.6	120	194.24
....		
3.2	640	36.16
3.3	660	25.76
3.4	680	15.04
3.5	700	4
3.6	720	-7.36

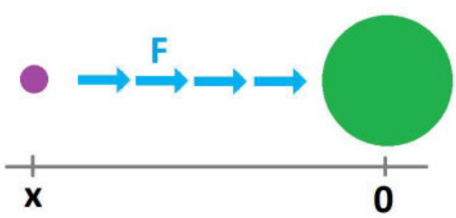


Example 6.7.10: "solar system"

The *gravity is constant* (independent of location) in the free-fall model:



If we now look at the solar system, will the gravity be constant too? Looking from the Earth, it certainly seems possible:



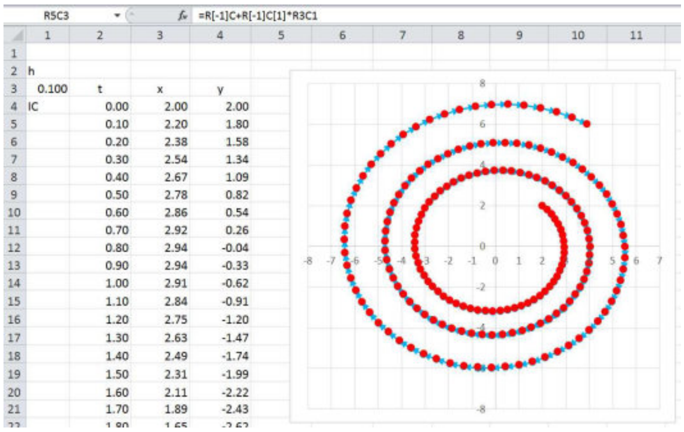
The direction will, of course, matter!

Exercise 6.7.11

Implement the model and examine the trajectories.

Exercise 6.7.12

Suggest a formula for the dependence of the force on the location that would produce such a spiral trajectory:

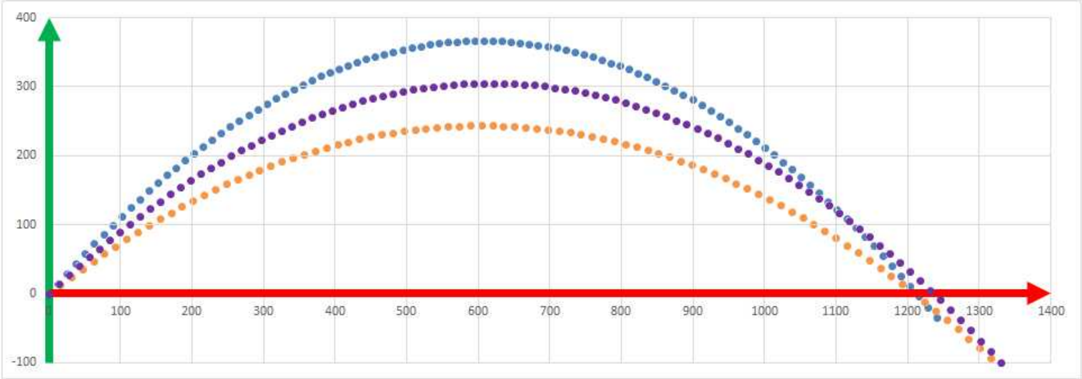


	1	2	3	4
1				
2	h			
3	0.100	t	x	y
4	IC	0.00	2.00	2.00
5		0.10	2.20	1.80
6		0.20	2.38	1.58
7		0.30	2.54	1.34
8		0.40	2.67	1.09
9		0.50	2.78	0.82
10		0.60	2.86	0.54
11		0.70	2.92	0.26
12		0.80	2.94	-0.04
13		0.90	2.94	-0.33
14		1.00	2.91	-0.62
15		1.10	2.84	-0.91
16		1.20	2.75	-1.20
17		1.30	2.63	-1.47
18		1.40	2.49	-1.74
19		1.50	2.31	-1.99
20		1.60	2.11	-2.22
21		1.70	1.89	-2.43

6.8. Optimization examples

Example 6.8.1: longest shot

In Chapter 3, we confirmed – numerically – the common knowledge fact that 45 degrees is the best angle to shoot for longer distance:



Let's apply the methods of differentiable calculus (Chapter 4) to *prove* this fact.

Recall the dynamics of this motion. The horizontal coordinate and the vertical coordinate are given as these functions of time t:

$$\begin{cases} x &= x_0 &+v_0 t \\ y &= y_0 &+u_0 t &-\dfrac{1}{2}gt^2 \end{cases}$$

They are subject to the following initial conditions:

- x_0 is the initial depth,
- v_0 is the initial horizontal component of velocity,
- y_0 is the initial height, and
- u_0 is the initial vertical component of velocity.

We shoot from the ground at zero elevation:

$$x_0 = y_0 = 0\,.$$

We also assume that the initial speed is 1 and that we shoot under angle α. Then, we have the specific

dynamics of this flight represented by two functions of time:

$$\begin{cases} v_0 &= \cos\alpha \\ u_0 &= \sin\alpha \end{cases}$$

We substitute:

$$\begin{cases} x &= \cos\alpha \cdot t \\ y &= \sin\alpha \cdot t \quad -\frac{1}{2}gt^2 \end{cases}$$

We will now express the length of the shot as the function of the angle α.

Now, the ground is reached at the moment t when $y = 0$. We, then, need to solve this system of equations:

$$\begin{cases} x &= \cos\alpha \cdot t \\ 0 &= \sin\alpha \cdot t \quad -\frac{1}{2}gt^2 \end{cases}$$

The second is a quadratic equation of t. We discard the starting moment, $t = 0$, and divide by $t \neq 0$:

$$\sin\alpha - \frac{1}{2}gt = 0\,.$$

We solve for t,

$$t = \frac{2}{g}\sin\alpha\,,$$

and substitute into the equation for x:

$$x = \cos\alpha \cdot t = \cos\alpha \cdot \frac{2}{g}\sin\alpha\,.$$

This is the final result, the depth of the shot as a function of the angle α:

$$\boxed{D(\alpha) = \frac{2}{g}\sin\alpha\cos\alpha}$$

We need to find the value of α that maximizes D. Since $D(0) = D(\pi/2) = 0$, the answer lies within this interval. We differentiate:

$$D'(\alpha) = \frac{2}{g}\big(\cos\alpha\cos\alpha + \sin\alpha(-\sin\alpha)\big) = \frac{2}{g}(\cos^2\alpha - \sin^2\alpha)\,.$$

We set it equal to 0 and conclude:

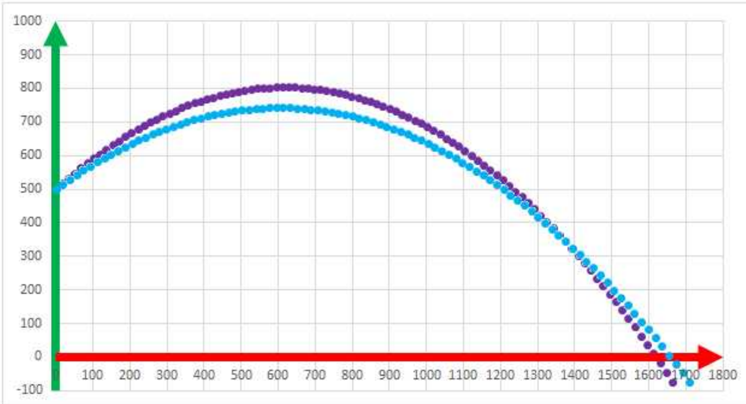
$$D'(\alpha) = 0 \implies \cos^2\alpha = \sin^2\alpha \implies \cos\alpha = \sin\alpha \implies \alpha = \frac{\pi}{4}\,.$$

Exercise 6.8.2

Use a trigonometric formula to finish the solution without differentiation.

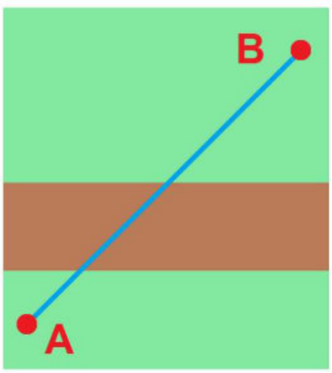
Exercise 6.8.3

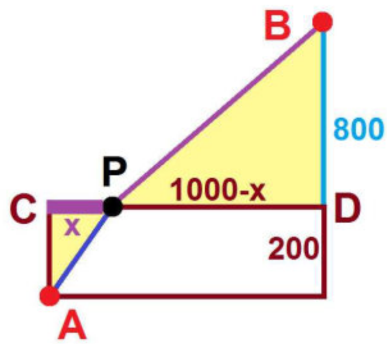
What if we shoot from a hill?



Example 6.8.4: laying a pipe

Suppose we are to lay a pipe from point A to point B located one kilometer east and one kilometer north of A. The cost of laying a pipe is normally \$50 per meter, except for a rocky strip of land 200 meters wide that goes east-west; here the price is \$100 per meter. What is the lowest cost to lay this pipe?





First we discover that it doesn't matter where the patch is located and assume that we have to cross it first. We proceed from A to point P on the other side of the patch and then to B. Then the cost C is

$$C = |AP| \cdot 100 + |PB| \cdot 50 \,.$$

We next denote by x the distance from P to the point directly across from A. Then

$$|AP| = \sqrt{200^2 + x^2} \quad \text{and} \quad |PB| = \sqrt{(1000 - x)^2 + 800^2} \,.$$

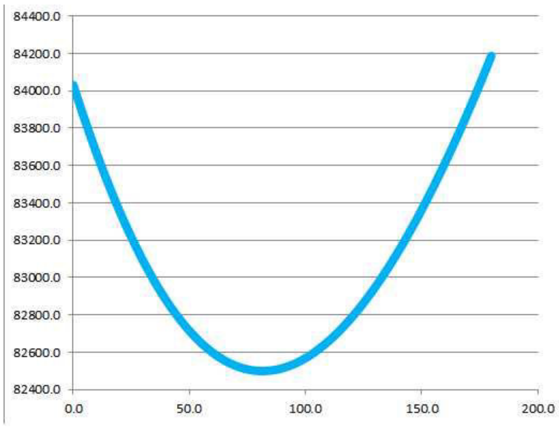
Therefore, we are to *minimize* the cost function:

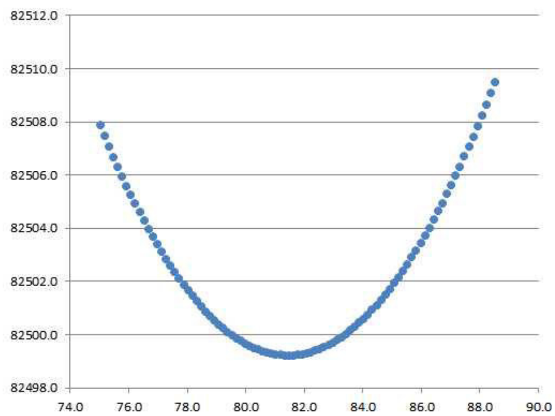
$$C(x) = \sqrt{200^2 + x^2} \cdot 100 + \sqrt{(1000 - x)^2 + 800^2} \cdot 50 \,.$$

Now, we convert this formula into a spreadsheet formula:

```
=SQRT(200^2+RC[-1]^2)*100+SQRT((1000-RC[-1])^2+800^2)*50
```

Plotting the curve, and then zooming in, suggests that the optimal choice for x is about 81.5 meters with the cost about \$82,498.50:





Let's confirm the result with calculus. Differentiate:

$$\begin{aligned} C'(x) &= \left(\sqrt{200^2+x^2}\cdot 100 + \sqrt{(1000-x)^2+800^2}\cdot 50\right)' \\ &= \frac{2x}{2\sqrt{200^2+x^2}}\cdot 100 + \frac{-2(1000-x)}{2\sqrt{(1000-x)^2+800^2}}\cdot 50\,. \end{aligned}$$

The equation $C'(x)=0$ proves itself too complex to be solved exactly.

Instead, we look back at the picture to recognize the terms in this expression as fractions of sides of these two triangles:

$$C'(x) = \cos\widehat{APC}\cdot 100 - \cos\widehat{BPD}\cdot 50\,.$$

Then, $C'=0$ if and only if

$$\frac{\cos\widehat{APC}}{\cos\widehat{BPD}} = \frac{50}{100}\,.$$

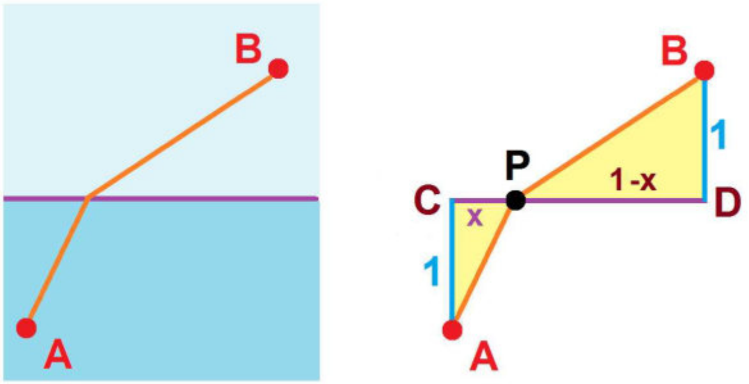
In other words, the optimal price is reached when the ratio of the cosines of the two angles at P are equal to the ratio of the two prices. In particular, making the price of laying the pipe across the patch more expensive will make this part of the path to cross more directly. Since cosine is a decreasing function on $(0,\pi/2)$, we conclude that $\widehat{BPD}$ is expressed uniquely in terms of $\widehat{APC}$.

Exercise 6.8.5

Show that, indeed, the location of the strip doesn't matter. Hint: Geometry.

Example 6.8.6: refraction

Suppose light is passing from one medium to another. It is known that the speed of light through the first is v_1, and through the second v_2. We rely on the principle that light follows the path of fastest speed and find the angle of refraction. A similar but simpler set-up:



Let x be the parameter, then the time it takes to get from A to B is:

$$\begin{aligned} T(x) &= \frac{|AP|}{v_1} + \frac{|AP|}{v_2} \\ &= \frac{\sqrt{1+x^2}}{v_1} + \frac{\sqrt{1+(1-x)^2}}{v_2}. \end{aligned}$$

We differentiate, just as in the last example:

$$\begin{aligned} T'(x) &= \frac{1}{v_1}\frac{x}{\sqrt{1+x^2}} + \frac{1}{v_2}\frac{-(1-x)}{\sqrt{1+(1-x)^2}} \\ &= \frac{1}{v_1}\cos\widehat{APC} - \frac{1}{v_2}\cos\widehat{BPD}. \end{aligned}$$

And $T' = 0$ if and only if

$$\frac{\cos\widehat{APC}}{\cos\widehat{BPD}} = \frac{v_1}{v_2}.$$

Therefore, light follows the path with the ratio of the cosines of the two angles at P equal to the ratio of the two propagation speeds.

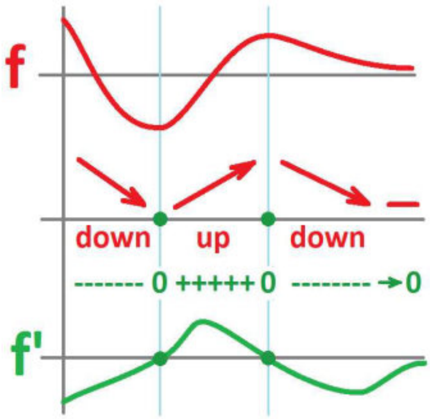
Exercise 6.8.7

Find the distance from the point $(1,1)$ to the parabola $y = -x^2$ by two methods: (a) find the minimal distance between the point and the curve, and (b) find the line perpendicular to the curve and its length.

Example 6.8.8: numerical optimization

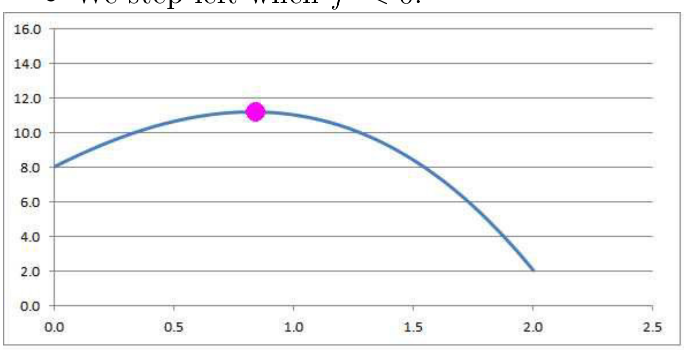
The differentiation might produce a function so complex that solving the equation $f' = 0$ analytically is impossible. In that case, we can apply one of the iterative processes of solving equation discussed earlier in this chapter.

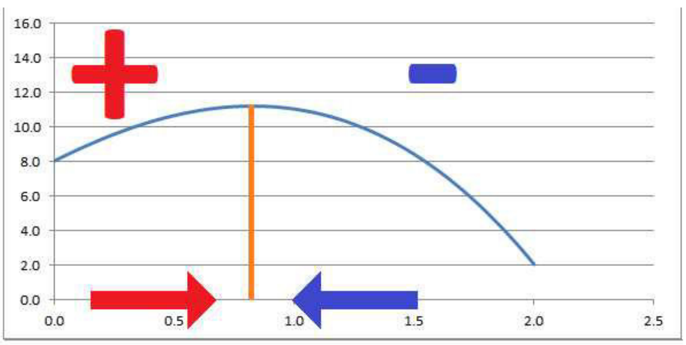
Alternatively, we design a process that follows the direction of the derivative as it always points at the nearest maximum:



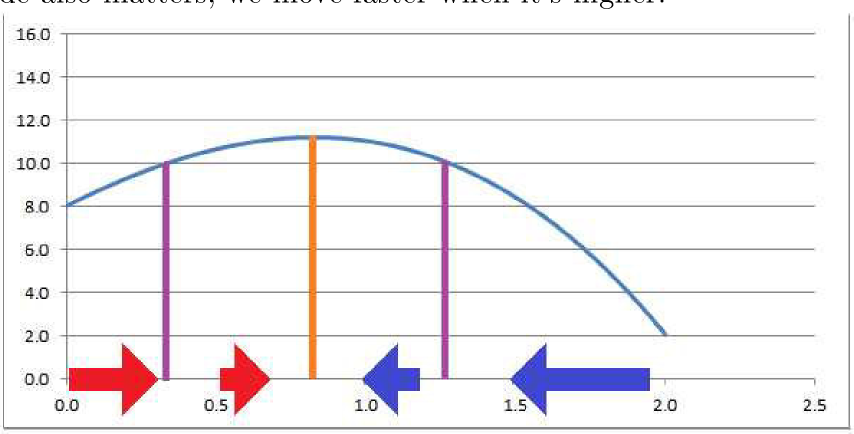
In other words, we move along the x-axis and then:

- We step right when $f' > 0$.
- We step left when $f' < 0$.





We reverse the direction when we look for a minimum. We make a step *proportional to the derivative* so that its magnitude also matters; we move faster when it's higher.



As you can see, the motion will slow down as we gets closer to the destination, which is convenient. So, we build a sequence of approximations *recursively*:

$$x_{n+1} = x_n + h \cdot f'(x_n)\,,$$

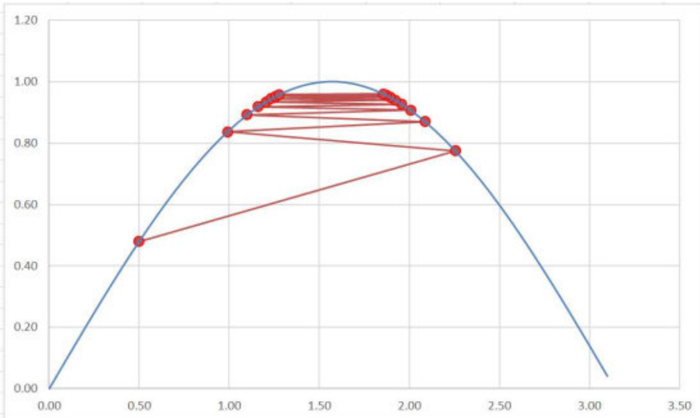
where h is the coefficient of proportionality. For example, for $f(x) = \sin x$, we have:

$$x_{n+1} = x_n + h \cdot \cos(x_n)\,.$$

The spreadsheet formula is:

```
=R[-1]C+R2C*COS(R[-1]C)
```

h=	2.00	
n	Xn	y=sin(Xn)
1	0.50	0.48
2	2.26	0.77
3	0.99	0.84
4	2.09	0.87
5	1.10	0.89
6	2.01	0.91
7	1.16	0.92
8	1.96	0.93
9	1.20	0.93
10	1.92	0.94
11	1.23	0.94
12	1.89	0.95
13	1.26	0.95
14	1.87	0.95
15	1.28	0.96
16	1.86	0.96



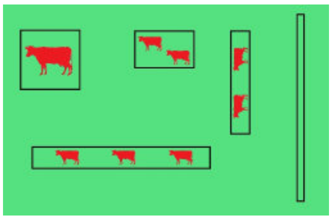
Especially for functions of several variables (Volume 4, Chapter 4HD-3), the method is called the *gradient descent* (which includes the ascent).

6.9. Functions of several variables

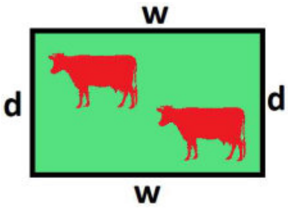
Example 6.9.1: cattle enclosure

Recall the problem (Chapter 2): A farmer with 100 yards of fencing material wants to build as large a rectangular enclosure as possible for his cattle. This is how the analysis starts.

The scope of possibilities is infinite:



We are supposed to find the dimensions of a rectangle – the width and the depth – with the largest area.



We start by randomly choosing possible measurements of the enclosure and compute its area with the formula:

$$\text{Area} = \text{width} \cdot \text{depth} .$$

We start with a rectangle with width 20 and increase the depth:

- 20 by 20 gives us an area of 400 square yards.
- 20 by 30 gives us an area of 600 square yards.
- 20 by 40 gives us an area of 800 square yards, etc.

Of course, the area is getting bigger and bigger; however, 30 by 30 gives us 900! The pattern is unclear. We will need to collect more data. Let's speed up this process with a spreadsheet. We introduce the variables:

- w is for the width, and
- d is for the depth, then

- a is for the area:

$$a = w \cdot d\,.$$

We first need to compile all possible combinations of the width (column W) and the depth (column D). We choose to go every 10 yards. Then the two quantities, independently, run through these 11 numbers:

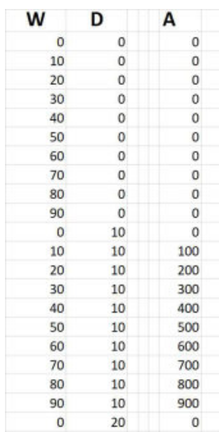
$$\text{width } = 0,\ 10,\ 20,\ ...,100 \ \text{ and depth } = 0,\ 10,\ 20,\ ...,100\,.$$

Together, there are $11 \cdot 11 = 121$ possible combinations.

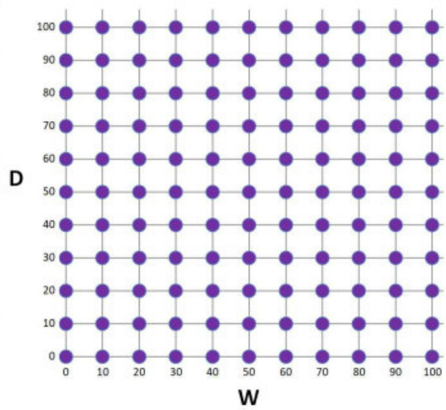
The first challenge is to list all possible pairs of width and depths in a spreadsheet. The simplest approach is to fix one value of w, starting with 0, and then start varying the value of d until we reach 10, then we set w equal to 10 and so on. Once we have them all, we also have all the areas too; we just compute the area, column A, with the formula:

```
=RC[-2]*RC[-1]
```

This is the result (left):



W	D	A
0	0	0
10	0	0
20	0	0
30	0	0
40	0	0
50	0	0
60	0	0
70	0	0
80	0	0
90	0	0
0	10	0
10	10	100
20	10	200
30	10	300
40	10	400
50	10	500
60	10	600
70	10	700
80	10	800
90	10	900
0	20	0



This list arrangement of the data is inconvenient. To investigate, let's *plot* these pairs (right). Together, they form an 11×11 *square* of possible combinations, with its width and depth corresponding to the width and depth of the enclosure! It appears that it is better to arrange these pairs in a *table* than in a list.

We choose to consider the dimensions every 10 yards via these two sets named after these two quantities:

$$W = \{0,\ 10,\ ...,100\} \quad \text{and} \quad D = \{0,\ 10,\ ...,100\}\,.$$

The table, and the spreadsheet, takes the form:

$W\backslash D$	0	10	...	100
0				
1				
$\vdots$				
100				

We put the data in a spreadsheet to compute the area of the enclosure with these dimensions according to the formula:

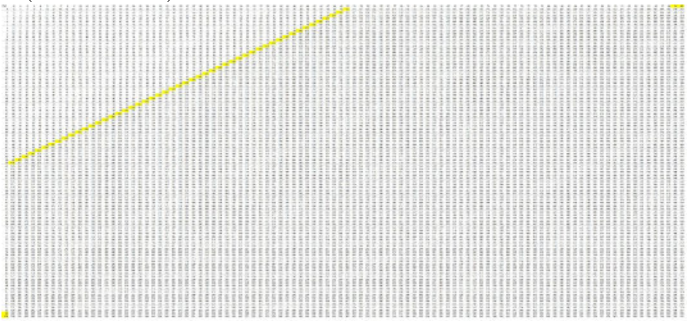
```
=RC2*R2C
```

referring to the same row and second column, and the second row and the same column, respectively. As a result, the table is filled with these values (shown for the 11×11 table):

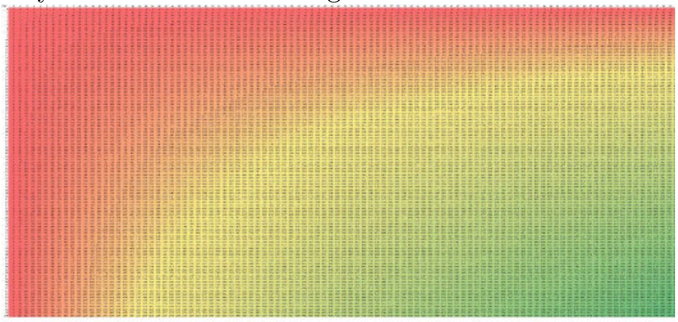
Areas

W\D	0	10	20	30	40	50	60	70	80	90	100
0	0	0	0	0	0	0	0	0	0	0	0
10	0	100	200	300	400	500	600	700	800	900	1000
20	0	200	400	600	800	1000	1200	1400	1600	1800	2000
30	0	300	600	900	1200	1500	1800	2100	2400	2700	3000
40	0	400	800	1200	1600	2000	2400	2800	3200	3600	4000
50	0	500	1000	1500	2000	2500	3000	3500	4000	4500	5000
60	0	600	1200	1800	2400	3000	3600	4200	4800	5400	6000
70	0	700	1400	2100	2800	3500	4200	4900	5600	6300	7000
80	0	800	1600	2400	3200	4000	4800	5600	6400	7200	8000
90	0	900	1800	2700	3600	4500	5400	6300	7200	8100	9000
100	0	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000

This is the complete table (101×101):



Now we color – automatically – the cells according to the value of the area it corresponds to:



The value of the area grows as we increase the width or the depth, but the fastest growth seems to be in some diagonal direction.

Any formula with two independent variables and one dependent variable can be studied in this manner:

$$a = wd \quad \text{or} \quad z = x + y\,,$$

Such an expression is called a *function of two variables*. The notation is as follows:

$$g(w,d) = wd \quad \text{or} \quad f(x,y) = x + y\,.$$

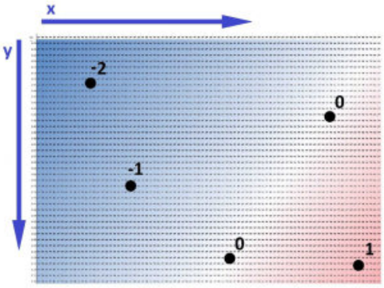
Example 6.9.2: function of two variables

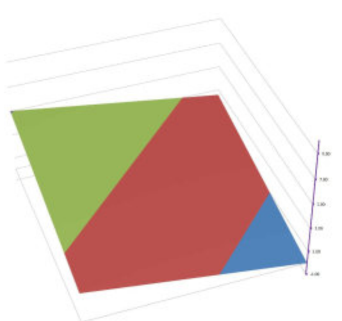
Let

$$f(x,y) = x + y\,.$$

We illustrate this new function below. First, by changing – independently — the two variables, we create a table of numbers (left). We can furthermore color this array of cells (middle) so that the color of the (x,y)-cell is determined by the value of z:

x\y	0	10	20	30	40	50	60
0	0	10	20	30	40	50	60
10	10	20	30	40	50	60	70
20	20	30	40	50	60	70	80
30	30	40	50	60	70	80	90
40	40	50	60	70	80	90	100
50	50	60	70	80	90	100	110
60	60	70	80	90	100	110	120
70	70	80	90	100	110	120	130
80	80	90	100	110	120	130	140





The value of z can also be visualized as the elevation of a point at that location (right).

So, the main metaphor for a function of two variables will be *terrain*:

Each line indicates a constant elevation.

Example 6.9.3: distance

The distance formula for the Cartesian plane (seen in Volume 1, Chapter 1PC-5) creates a function of two variables. This is the distance from a point (x, y) to the origin::

$$z = \sqrt{x^2 + y^2}\,.$$

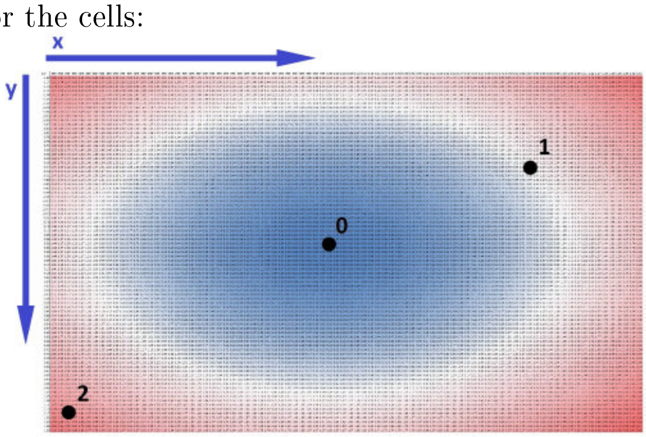
Slightly simpler is the square of the distance from a point (x, y) to the origin:

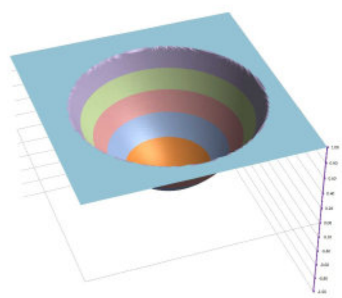
$$z = x^2 + y^2\,.$$

We create a table of the values of the expression on the left in a spreadsheet with the formula:

```
=RC1^2+R1C^2
```

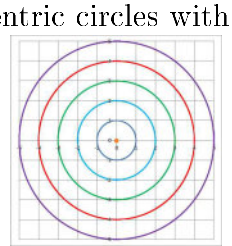
We then color the cells:





The negative values of z are in blue and the positive are in red. The circular pattern is clear.

The pattern seems to be made from concentric circles with the radius that varies with z:



For each z, we have a *relation* between x and y.

We also represent a function p diagrammatically as a *black box* that processes the inputs and produces the

output:

$$\begin{array}{cccc} \text{inputs} & & \text{function} & & \text{output} \\ x & & & & \\ & \searrow & & & \\ & & \boxed{p} & \to & z \\ & \nearrow & & & \\ y & & & & \end{array}$$

Instead, we would like to see a single input variable, (x, y), decomposed into two x and y to be processed by the function *at the same time*:

$$(x, y) \quad \to \quad \boxed{p} \quad \to \quad z$$

The difference from all the functions we have seen until now is the nature of the input.

So, even though we speak of *two* variables, the idea of function remains the same:

> ▶ There is a set (domain) and another set (codomain), and the function assigns to each element of the former an element of the latter.

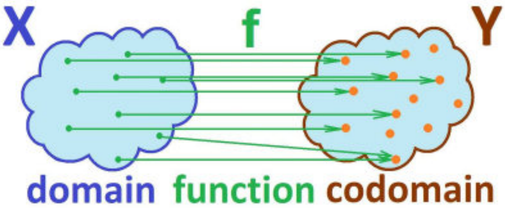
The idea is reflected in the notation we use:

$$F : X \to Z$$

or

$$X \xrightarrow{F} Z$$

A common way to visualize the concept of function – especially when the sets cannot be represented by mere lists – is to draw shapeless blobs connected by arrows:



In contrast to numerical function, however, *the domain is a subset of the* (x, y)*-plane.*

For example, we have for $f(x, y) = x + y$:

$$(0, 0) \to 0,\ (0, 1) \to 1,\ (1, 0) \to 1,\ (1, 1) \to 2,\ (1, 2) \to 3,\ (2, 1) \to 3\ldots$$

Each arrow clearly identifies the *input* – an element of X – of this procedure by its beginning, and the *output* – an element of Z – by its end.

This is the notation for the output of a function F when the input is x:

Input and output of function

$$F(x, y) = z$$

or

$$F : (x, y) \quad \to \quad z$$

It reads "F of (x, y) is z".

We still have:

$$F(\text{ input }) = \text{ output}$$

and

$$F : \text{ input } \quad \to \quad \text{ output }.$$

Functions are *explicit relations*. There are *three* variables related to each other, but this relation is unequal: The two input variables come first and, therefore, the output is *dependent* on the input. That is why we say that the inputs are the *independent variables* while the output is the *dependent variable*.

Example 6.9.4: flowcharts represent functions

For example, for a given input (x, y), we first split it: x and y are the two *numerical* inputs. Then we do the following consecutively:

- add x and y,
- multiply by 2, and then
- square.

Such a procedure can be conveniently visualized with a "flowchart":

$$(x,y) \to \boxed{x+y} \to u \to \boxed{u\cdot 2} \to z \to \boxed{z^2} \to v$$

Functions of two variables come from many sources and can be expressed in different forms:

- a list of instructions (an algorithm)
- an algebraic formula
- a list of pairs of inputs and outputs
- a graph
- a transformation

An *algebraic representation* is exemplified by $z = x^2y$. In order to properly introduce this as a function, we give it a name, say f, and write:

$$f(x, y) = x^2y\,.$$

Let's examine this notation:

Function of two variables

	z	$=$	f	$($	$x\ ,\ y$	$)$	$=$	x^2y
	$\uparrow$		$\uparrow$		$\uparrow\ \uparrow$			$\uparrow\uparrow$
name:	dependent variable		function		independent variables			independent variables

Example 6.9.5: plug in values

Insert one input value in all of these boxes and the other in those circles. For example, this function:

$$f(x) = \frac{2x^2y - 3y + 7}{y^3 + 2x + 1}\,,$$

can be understood and evaluated via this diagram:

$$f(\square) = \frac{2\square^2 \bigcirc -3 \bigcirc +7}{\bigcirc^3 + 2\square + 1}\,.$$

This is how $f(3, 0)$ is evaluated:

$$f\left(\boxed{3}, 0\right) = \frac{2\boxed{3}^2 0 - 3 0 + 7}{0^3 + 2\boxed{3} + 1}\,.$$

In summary,

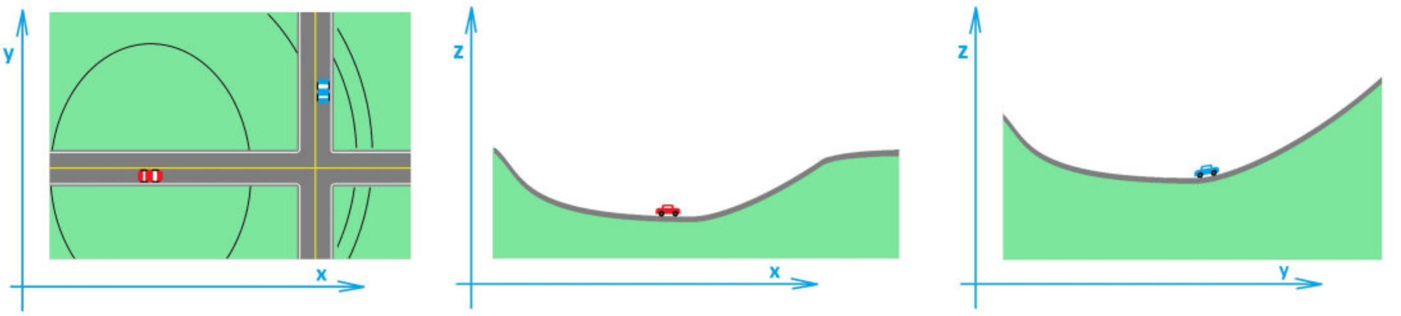
▶ "x" and "y" in a formula serve as a *placeholders* for: numbers, variables, and whole functions.

How do we study a function of two variables? We use what we know about functions of single variable.

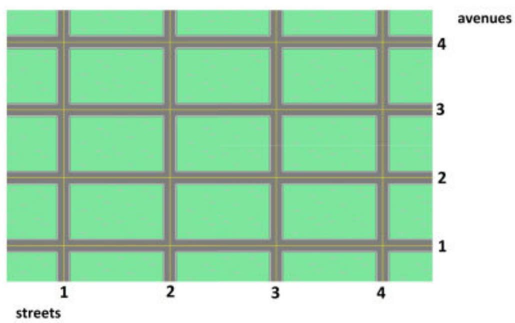
Above we looked at the curves of constant elevation of the surfaces. An alternative idea is a surveying method:

▶ In order to study a terrain, we concentrate on the two main directions.

Imagine that we drive south-north and east-west separately watching how the elevation changes:

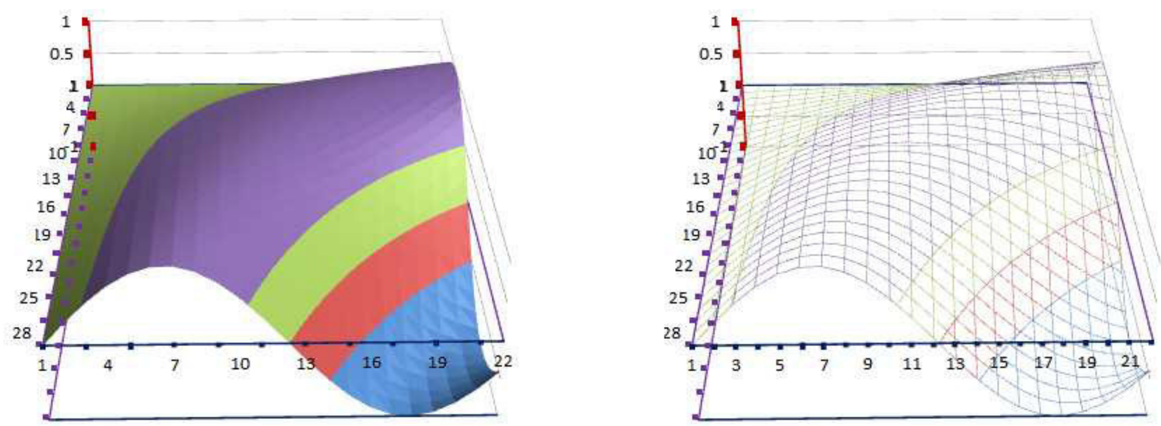


We can even imagine that we drive around a city on a hill and these trips follow the street grid:



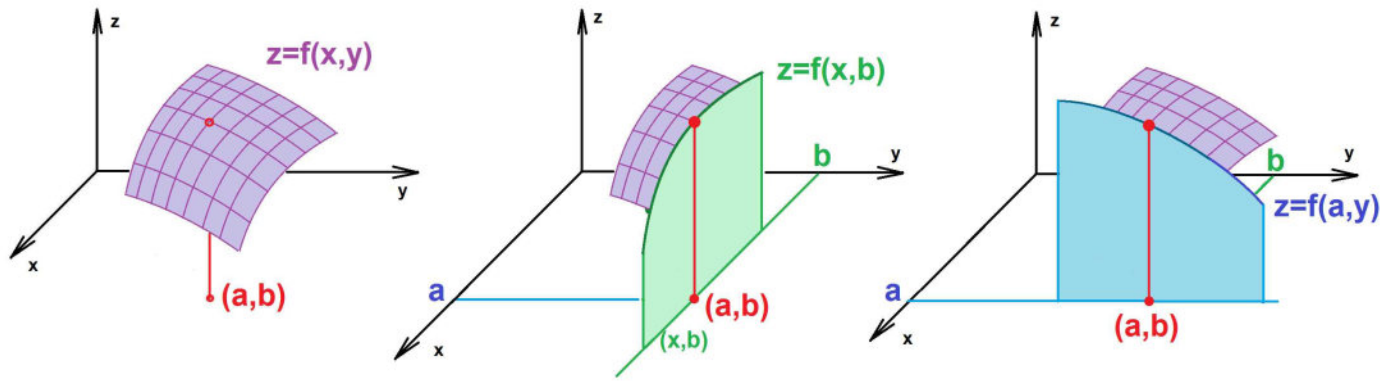
Each of these trips creates a function of single variable, x or y.

To visualize, consider the plot of $F(x, y) = \sin(xy)$ on the left:



We plot the surface as a "wire-frame" on the right. Each wire is a separate trip.

The graphs of these functions are the slices cut by the vertical planes aligned with the axes from the surface that is the graph of F:

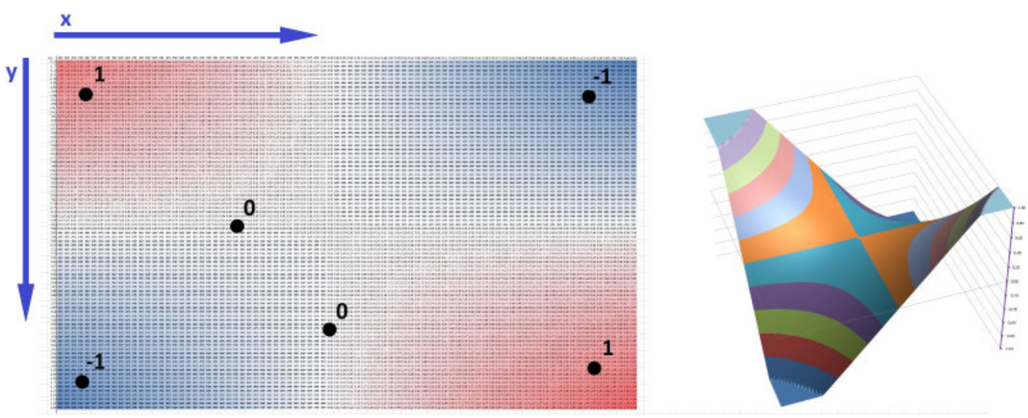


As a first step, we can study the monotonicity of these functions with the help of their derivatives.

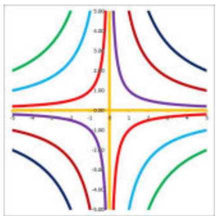
Example 6.9.6: saddle

Let's plot the graph of the function:

$$z = xy\,.$$



This is what the graphs of these relations look like plotted for various z's; they are curves called *hyperbolas*:

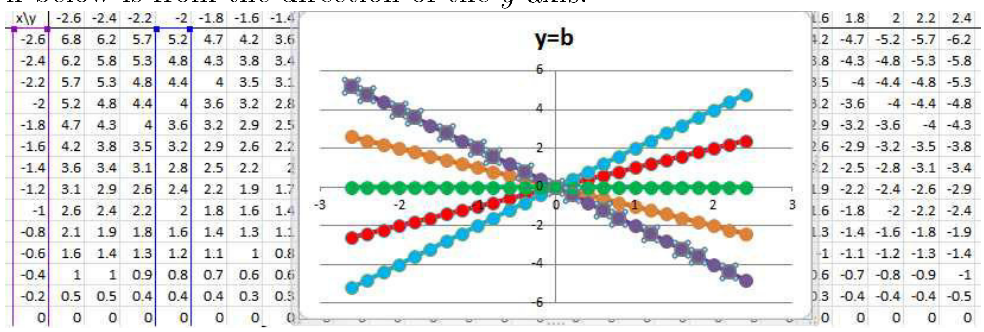


Instead, we fix one independent variable at a time.

We fix y first:

plane	equation	curve
$y = 2$	$z = x \cdot 2$	line with slope 2
$y = 1$	$z = x \cdot 1$	line with slope 1
$y = 0$	$z = x \cdot 0 = 0$	line with slope 0
$y = -1$	$z = x \cdot (-1)$	line with slope 1
$y = -2$	$z = x \cdot (-2)$	line with slope -2

The view shown below is from the direction of the y-axis:

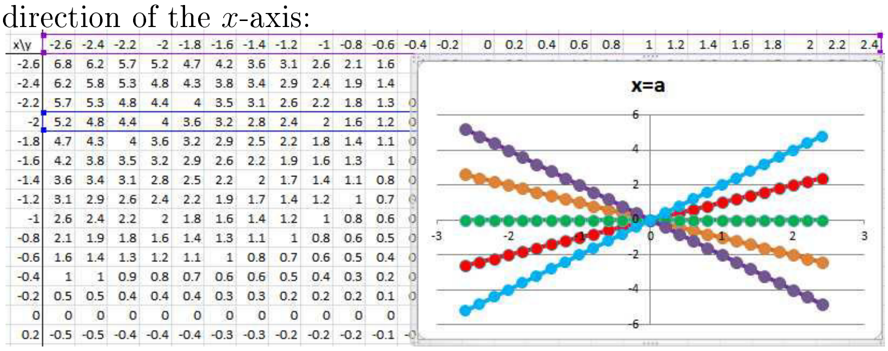


The data for each line comes from the x-column of the spreadsheet and one of the z-columns. These lines give the lines of elevation of this terrain in a particular, say, east-west direction. This is equivalent to cutting the graph by a vertical plane parallel to the xz-plane.

We fix x second:

plane	equation	curve
$x = 2$	$z = 2 \cdot y$	line with slope 2
$x = 1$	$z = 1 \cdot y$	line with slope 1
$x = 0$	$z = 0 \cdot y = 0$	line with slope 0
$x = -1$	$z = (-1) \cdot y$	line with slope 1
$x = -2$	$z = (-2) \cdot y$	line with slope -2

This is equivalent to cutting the graph by a vertical plane parallel to the yz-plane. The view shown below is from the direction of the x-axis:



x\y	-2.6	-2.4	-2.2	-2	-1.8	-1.6	-1.4	-1.2	-1	-0.8	-0.6	-0.4	-0.2	0	0.2	0.4	0.6	0.8	1	1.2	1.4	1.6	1.8	2	2.2	2.4
-2.6	6.8	6.2	5.7	5.2	4.7	4.2	3.6	3.1	2.6	2.1	1.6															
-2.4	6.2	5.8	5.3	4.8	4.3	3.8	3.4	2.9	2.4	1.9	1.4															
-2.2	5.7	5.3	4.8	4.4	4	3.5	3.1	2.6	2.2	1.8	1.3															
-2	5.2	4.8	4.4	4	3.6	3.2	2.8	2.4	2	1.6	1.2															
-1.8	4.7	4.3	4	3.6	3.2	2.9	2.5	2.2	1.8	1.4	1.1															
-1.6	4.2	3.8	3.5	3.2	2.9	2.6	2.2	1.9	1.6	1.3	1															
-1.4	3.6	3.4	3.1	2.8	2.5	2.2	2	1.7	1.4	1.1	0.8															
-1.2	3.1	2.9	2.6	2.4	2.2	1.9	1.7	1.4	1.2	1	0.7															
-1	2.6	2.4	2.2	2	1.8	1.6	1.4	1.2	1	0.8	0.6															
-0.8	2.1	1.9	1.8	1.6	1.4	1.3	1.1	1	0.8	0.6	0.5															
-0.6	1.6	1.4	1.3	1.2	1.1	1	0.8	0.7	0.6	0.5	0.4															
-0.4	1	1	0.9	0.8	0.7	0.6	0.6	0.5	0.4	0.3	0.2															
-0.2	0.5	0.5	0.4	0.4	0.4	0.3	0.3	0.2	0.2	0.2	0.1															
0	0	0	0	0	0	0	0	0	0	0	0															
0.2	-0.5	-0.5	-0.4	-0.4	-0.4	-0.3	-0.3	-0.2	-0.2	-0.2	-0.1															

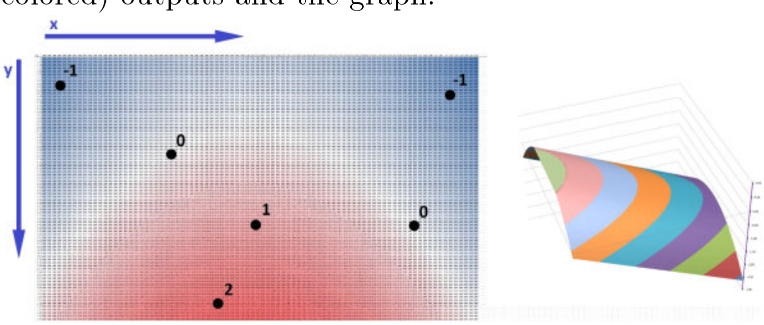
The data for each line comes from the y-row of the spreadsheet and one of the z-rows. These lines give the lines of elevation of this terrain in a particular, say, north-south direction.

Example 6.9.7: another surface

Let's try a similar analysis for the function below:

$$z = y - x^2 \,.$$

Below is the array of (colored) outputs and the graph:



If y is fixed at $y = b$, we have a function of x:

$$b - x^2 \,.$$

Its derivative with respect to x is $-2x$. Therefore, if we move along the x-axis, we go up and then down no matter where this line is. It's a parabola!

If x is fixed at $x = a$, we have a function of y:

$$y - a^2 \,.$$

Its derivative with respect to y is 1. Therefore, if we move along the y-axis, we go up no matter where this line is. It's a straight line!

Exercise 6.9.8

1. Sketch a straight line with parabolas hanging from it.
2. Sketch a parabola with straight lines leaned against it.

Exercise 6.9.9

Provide a similar analysis for $f(x, y) = x^2 + y^2$.

This study will continue in Volume 3, Chapter 3IC-4, and then in Volume 4, Chapter 4HD-3.

Exercises

Contents

1. Exercises: Sets, logic, functions

Exercise 1.1

Represent the following set in the set-building notation:

$$X = [0,1] \cup [2,3] = \,?$$

Exercise 1.2

Simplify:

$$\{x > 0 : \ x \text{ is a negative integer }\}\,.$$

Exercise 1.3

What are the max, min, and any bounds of the set of integers? What about $\mathbf{R}$?

Exercise 1.4

Is the converse of a true statement true?

Exercise 1.5

Is the converse of the converse of a true statement true?

Exercise 1.6

State the converse of this statement: "The converse of the converse of a true statement is true".

Exercise 1.7

Represent these sets as intersections and unions:

1. $(0,5)$
2. $\{3\}$
3. $\varnothing$
4. $\{x : x > 0 \ \texttt{OR} \ \ x \ \text{ is an integer}\}$
5. $\{x : \ x \ \text{ is divisible by } \ 6\}$

Exercise 1.8

True or false: $0 = 1 \implies 0 = 1$?

Exercise 1.9

Prove or disprove:

$$\max\{\max A, \max B\} = \max(A \cup B)\,.$$

Exercise 1.10

(a) If, starting with a statement A, after a series of conclusions you arrive to $0 = 1$, what can you conclude about A? (b) If, starting with a statement A, after a series of conclusions you arrive to $0 = 0$, what can you conclude about A?

Exercise 1.11

We know that "If it rains, the road gets wet". Does it mean that if the road is wet, it has rained?

Exercise 1.12

A garage light is controlled by a switch and, also, it may automatically turn on when it senses motion during nighttime. If the light is OFF, what do you conclude?

Exercise 1.13

If an advertisement claims that "All our second-hand cars come with working AC", what is the easiest way to disprove the sentence?

Exercise 1.14

Teachers often say to the student's parents: "If your student works harder, he'll improve". When he won't improve and the parents come back to the teacher, he will answer: "He didn't improve, that means he didn't work harder". Analyze.

Exercise 1.15

Give the definition of a circle.

Exercise 1.16

Suppose the cost is $f(x)$ dollars for a taxi trip of x miles. Interpret the following stories in terms of f.

1. Monday, I took a taxi to the station 5 miles away.
2. Tuesday, I took a taxi to the station but then realized that I left something at home and had to come back.
3. Wednesday, I took a taxi to the station and I gave my driver a five dollar tip.
4. Thursday, I took a taxi to the station but the driver got lost and drove five extra miles.
5. Friday, I have been taking a taxi to the station all week on credit; I pay what I owe today.

What if there is an extra charge per ride of m dollars?

Exercise 1.17

Let $f : A \to B$ and $g : C \to D$ be two possible functions. For each of the following functions, state whether or not you can compute $f \circ g$:

- $D \subset B$
- $C \subset A$
- $B \subset D$
- $B = C$

Exercise 1.18

Function $y = f(x)$ is given below by a list of some of its values. Make sure the function is onto.

x	-1	0	1	2	3	4	5
$y = f(x)$	-1		4	5		2	

Exercise 1.19

Function $y = f(x)$ is given below by a list of some of its values. Add missing values in such a way that the function is one-to-one.

x	-1	0	1	2	3	4	5
$y = f(x)$	-1		0	5		0	

Exercise 1.20

Represent the function $h(x) = \sin^2 x + \sin^3 x$ as the composition $g \circ f$ of two functions $y = f(x)$

and $z = g(y)$.

Exercise 1.21

Function $y = f(x)$ is given below by a list its values. Find its inverse and represent it by a similar table.

x	0	1	2	3	4
$y = f(x)$	0	1	2	4	3

Exercise 1.22

Find the formulas of the inverses of the following functions: (a) $f(x) = (x+1)^3$; (b) $g(x) = \ln(x^3)$.

Exercise 1.23

Are the following functions invertible? (a) $y = f(n)$ is the number of students in your class whose birthday is on the nth day of the year. (b) $y = f(t)$ is the total accumulated rainfall in inches t on a given day in a particular location.

Exercise 1.24

Given the tables of values of f, g, find the table of values of $f \circ g$:

x	$y = g(x)$
0	0
1	4
2	3
3	0
4	1

y	$z = f(y)$
0	4
1	4
2	0
3	1
4	2

What if the last rows were missing?

Exercise 1.25

Represent the function below as the composition $f \circ g$ of two functions:

$$h(x) = \sqrt{2x^3 + x}\,.$$

Exercise 1.26

Represent the function $h(x) = 2\sin^3 x + \sin x + 5$ as the composition of two functions one of which is trigonometric.

Exercise 1.27

(a) Represent the function $h(x) = e^{x^3-1}$, as the composition of two functions f and g. (b) Provide formulas for the two possible compositions of the two functions: "take the logarithm base 2 of" and "take the square root of".

Exercise 1.28

Suppose a function f performs the operation: "take the logarithm base 2 of", and function g performs: "take the square root of". (a) Verbally describe the inverses of f and g. (b) Find the formulas for these four functions. (c) Give them domains and codomains.

Exercise 1.29

1. Represent the function $h(x) = \sqrt{x^2 - 1}$ as the composition of two functions f and g.
2. Provide a formula for the composition $y = f(g(x))$ of $f(u) = u^2 + u$ and $g(x) = 2x - 1$.

Exercise 1.30

Provide a formula for the composition $y = f(g(x))$ of $f(u) = \sin u$ and $g(x) = \sqrt{x}$.

Exercise 1.31

Provide a formula for the composition $y = f(g(x))$ of $f(u) = u^2 - 3u + 2$ and $g(x) = x$.

Exercise 1.32

What is the meaning of the inverse of the function $f(x) = 3x^2 + 1$? Hint: Choose appropriate domains.

Exercise 1.33

1. Represent the function $h(x) = \sqrt{x - 1}$ as the composition of two functions.
2. Represent the function $k(t) = \sqrt{t^2 - 1}$ as the composition of three functions.
3. Represent the function $p(t) = \sin\sqrt{t^2 - 1}$ as the composition of four functions.

Exercise 1.34

(a) What is the composition $f \circ g$ for the functions given by $f(u) = u^2 + u$ and $g(x) = 3$? (a) What is the composition $f \circ g$ for the functions given by $f(u) = 2$ and $g(x) = \sqrt{x}$?

Exercise 1.35

Function $y = f(x)$ is given below by a list its values. Find its inverse and represent it by a similar table.

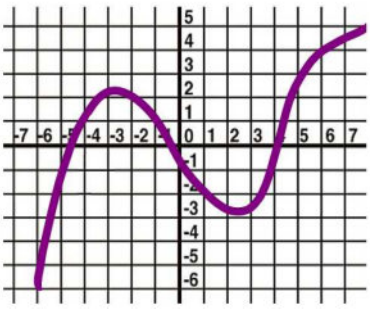
x	0	1	2	3	4
$y = f(x)$	1	2	0	4	3

Exercise 1.36

Give examples of functions that are their own inverses.

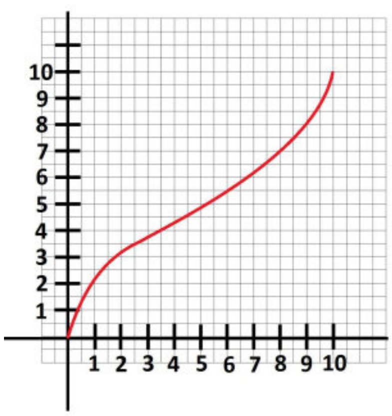
Exercise 1.37

Plot the inverse of the function shown below, if possible:



Exercise 1.38

Plot the graph of the inverse of this function, if possible:



Exercise 1.39

Represent this function: $h(x) = \tan(2x)$ as the composition of two functions of variables x and y.

Exercise 1.40

Represent this function: $h(x) = \dfrac{x^3+1}{x^3-1}$, as the composition of two functions of variables x and y.

Exercise 1.41

Function $y = f(x)$ is given below by a list of its values. Is the function one-to one? What about its inverse?

x	0	1	2	3	4
$y = f(x)$	0	1	2	1	2

Exercise 1.42

Function $y = f(x)$ is given below by a list of its values. Is the function one-to one? What about its inverse?

x	0	1	2	3	4
$y = f(x)$	7	5	3	4	6

Exercise 1.43

Functions $y = f(x)$ and $u = g(y)$ are given below by tables of some of their values. Present the composition $u = h(x)$ of these functions by a similar table:

x	0	1	2	3	4
$y = f(x)$	1	1	2	0	2

y	0	1	2	3	4
$u = g(y)$	3	1	2	1	0

Exercise 1.44

Function $y = f(x)$ is given below by a list of some of its values. Add missing values in such a way that the function is one-to one.

x	-1	0	1	2	3	4	5
$y = f(x)$	-1		4	5		2	

Exercise 1.45

Plot the graph of the function $f(x) = \dfrac{1}{x-1}$ and the graph of its inverse. Identify its important features.

Exercise 1.46

(a) Algebraically, show that the function $f(x) = x^2$ is not one-to-one. (b) Graphically, show that the function $g(x) = 2^{x+1}$ is one-to-one. (c) Find the inverse of g.

Exercise 1.47

Describe – both geometrically and algebraically – two different transformations that make a 1×1 square into a 2×3 rectangle.

2. Exercises: Background

Exercise 2.1

Find the equation of the line passing through the points $(-1, 1)$ and $(-1, 5)$.

Exercise 2.2

For the points $P = (0, 1)$, $Q = (1, 2)$, and $R = (-1, 2)$, determine the points that are symmetric with respect to the axis and the origin.

Exercise 2.3

Consider triangle ABC in the plane where $A = (3, 2)$, $B = (3, -3)$, $C = (-2, -2)$. Find the lengths of the sides of the triangle.

Exercise 2.4

Find all x such that the distance between the points $(3, -8)$ and $(x, -6)$ is 5.

Exercise 2.5

Find the perimeter of the triangle with the vertices at $(3, -1)$, $(3, 6)$, and $(-6, -5)$.

Exercise 2.6

Find the point on the x-axis that is equidistant from the points $(-1, 5)$ and $(6, 4)$.

Exercise 2.7

Find the distance between the points of intersection of the circle $(x-1)^2 + (y-2)^2 = 6$ with the axes.

Exercise 2.8

Solve the system of linear equations:

$$\begin{cases} x - y = -1, \\ 2x + y = 0. \end{cases}$$

Exercise 2.9

A movie theater charges \$10 for adults and \$6 for children. On a particular day when 320 people paid an admission, the total receipts were \$3120. How many were adults and how many were children?

Exercise 2.10

The taxi charges \$1.75 for the first quarter of a mile and \$0.35 for each additional fifth of a mile. Find a linear function which models the taxi fare f as a function of the number of miles driven, x.

Exercise 2.11

Find the value of k so that the line containing the points $(-6, 0)$ and $(k, -5)$ is parallel to the line containing the points $(4, 3)$ and $(1, 7)$.

Exercise 2.12

Is this a parabola?

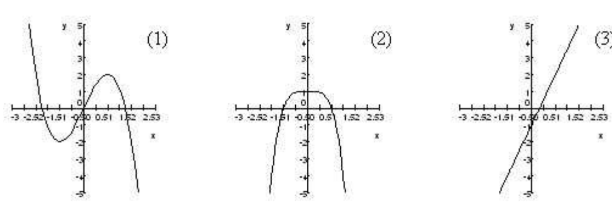
Exercise 2.13

Find the equation satisfied by all points that lie 2 units away from the point $(-1, -2)$ and by no other points.

Exercise 2.14

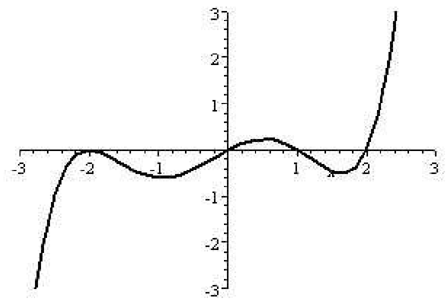
For the polynomials graphed below, find the following:

	1	2	3
smallest possible degree			
sign of the leading coefficient			
degree is odd/even			



Exercise 2.15

Find a possible formula for the function plotted below:



Exercise 2.16

A factory is to be built on a lot measuring 240 ft by 320 ft. A building code requires that a lawn of uniform width and equal in area to the factory must surround the factory. What must the width of the lawn be?

Exercise 2.17

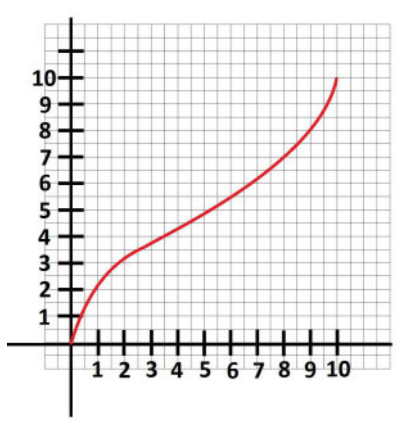
A factory occupies a lot measuring 240 ft by 320 ft. A building code requires that a lawn of uniform width and equal in area to the factory must surround the factory. What must the width of the lawn be?

Exercise 2.18

(a) Solve the equation $(x^2+1)(x+1)(x-1)=0$.
(b) Solve the inequality $(x^2+1)(x+1)(x-1)>0$.

Exercise 2.19

The graph of the function $y=f(x)$ is given below. (a) Find such a y that the point $(2,y)$ belongs to the graph. (b) Find such an x that the point $(x,3)$ belongs to the graph. (c) Find such an x that the point (x,x) belongs to the graph. Show your drawing.



Exercise 2.20

Make a flowchart and then provide a formula for the function $y=f(x)$ that represents a parking fee for a stay of x hours. It is computed as follows: free for the first hour and \$1 per hour beyond.

Exercise 2.21

Find the implied domain of the function:

$$\frac{x-1}{x+1}\ln(x^2+1)\sin x\,.$$

Exercise 2.22

Find the implied domain of the function:

$$(x-1)(x^2+1)2^x\,.$$

Exercise 2.23

Finish the sentence: "If a function fails the horizontal line test, then..."

Exercise 2.24

Explain the difference between these two functions:

$$\sqrt{\frac{x-1}{x+1}} \quad\text{and}\quad \frac{\sqrt{x-1}}{\sqrt{x+1}}\,.$$

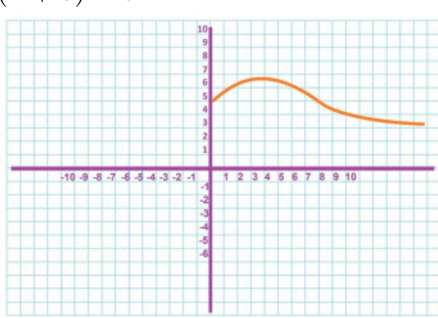
Exercise 2.25

Classify these functions:

function	odd	even	onto	one-to-one
$f(x)=2x-1$				
$g(x)=-x+2$				
$h(x)=3$				

Exercise 2.26

The graph of $y=f(x)$ is plotted below. Sketch $y=-f(x+5)-6$.



Exercise 2.27

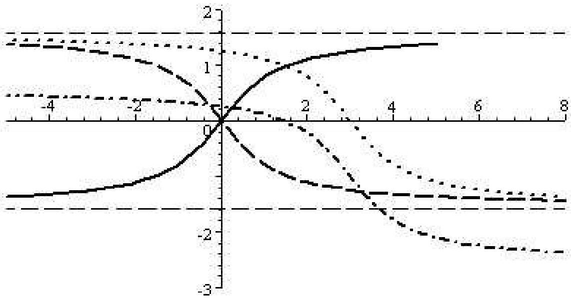
Is the composition of two functions that are odd-/even odd/even?

Exercise 2.28

Find the formulas of the inverses of the following functions: (a) $f(x)=(x+1)^3$; (b) $g(x)=\ln(x^3)$.

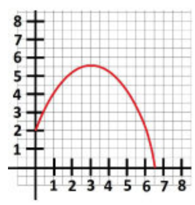
Exercise 2.29

One of the graphs below is that of $y = \arctan x$. What are the others?



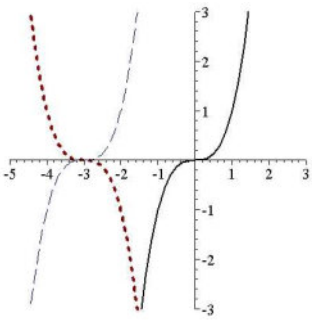
Exercise 2.30

The graph of the function $y = f(x)$ is given below. Sketch the graph of $y = 2f(3x)$ and then $y = f(-x) - 1$.



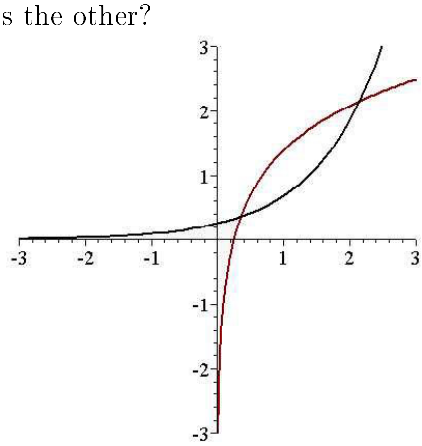
Exercise 2.31

The graph drawn with a solid line is $y = x^3$. What are the other two?



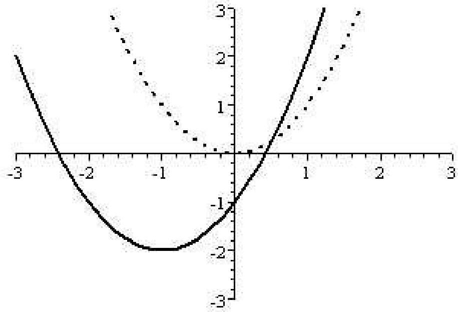
Exercise 2.32

The graph of one of the functions below is $y = e^x$. What is the other?



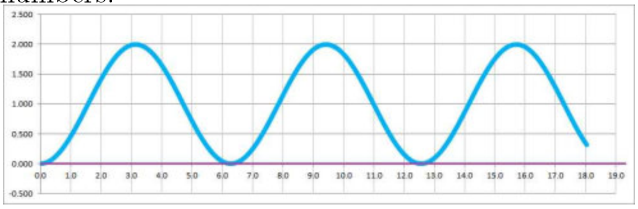
Exercise 2.33

The graphs below are parabolas. One is $y = x^2$. What is the other?



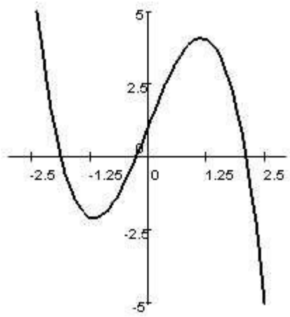
Exercise 2.34

The graph below is the graph of the function $f(x) = A \sin x + B$ for some A and B. Find these numbers.



Exercise 2.35

The graph of function f is given below. Sketch the graph of $y = 2f(x + 2) + 2$. Explain how you get it.

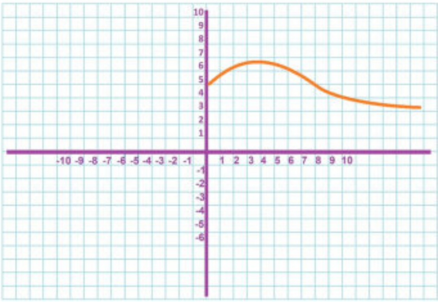


Exercise 2.36

By transforming the graph of $y = e^x$, plot the graph of the function $f(x) = 2e^{x-3}$. Identify the domain, the range, and the asymptotes.

Exercise 2.37

Half of the graph of an even function is shown below; provide the other half:

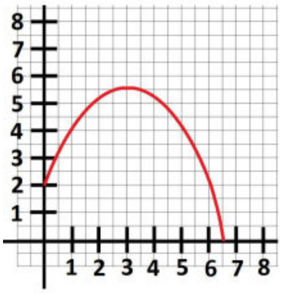


Exercise 2.38

Half of the graph of an odd function is shown above; provide the other half.

Exercise 2.39

The graph of the function $y = f(x)$ is given below. Sketch the graph of $y = 2f(x)$ and then $y = 2f(x) - 1$.

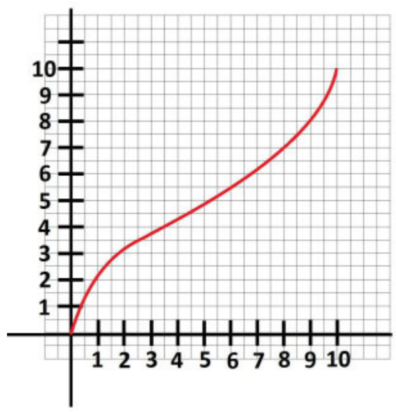


Exercise 2.40

The graph above is a parabola. Find its formula.

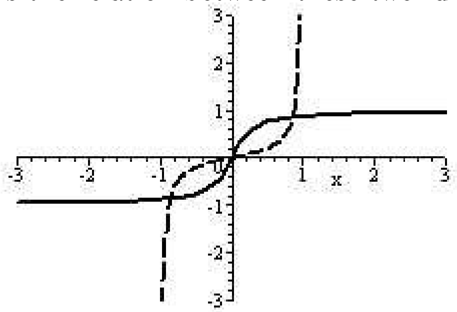
Exercise 2.41

The graph of the function $y = f(x)$ is given below. Sketch the graph of $y = \frac{1}{2}f(x)$ and then $y = \frac{1}{2}f(x-1)$.



Exercise 2.42

What is the relation between these two functions?



Exercise 2.43

Plot the graph of a function that is both odd and even.

Exercise 2.44

Give examples of odd and even functions that aren't polynomials.

Exercise 2.45

Is the inverse of an odd/even function odd/even?

Exercise 2.46

By transforming the graph of $y = \sin x$, plot the graph of the function $f(x) = 2\sin(x-3)$. Identify its range.

Exercise 2.47

Give examples of an even function, an odd function, and a function that's neither. Provide formulas.

3. Exercises: Sequences

Exercise 3.1

Present the first 5 terms of the sequence:

$$a_1 = 1, \quad a_{n+1} = -(a_n + 1).$$

Exercise 3.2

Represent in sigma notation:

$$-1 - 2 - 3 - 4 - 5 - \ldots - 10.$$

Exercise 3.3

Find the following sum:

$$-1 - 2 - 3 - 4 - 5 - \ldots - 10.$$

Exercise 3.4

Find the sequence of sums of the following sequence:

$$-1, 2, -4, 8, -5, \ldots$$

Exercise 3.5

Compute $\displaystyle\sum_{n=1}^{4} n^2$.

Exercise 3.6

Show that $\dfrac{n}{n+1}$ is an increasing sequence. What kind of sequence is $\dfrac{n+1}{n}$? Give examples of increasing and decreasing sequences.

Exercise 3.7

Find the next item in each list:

1. $7, 14, 28, 56, 112, \ldots$
2. $15, 27, 39, 51, 63, \ldots$
3. $197, 181, 165, 149, 133, \ldots$

Exercise 3.8

A pile of logs has 50 logs in the bottom layer, 49 logs in the next layer, 48 logs in the next layer, and so on, until the top layer has 1 log. How many logs are in the pile?

Exercise 3.9

In the beginning of each year, a person puts \$5000 in a bank that pays 3% compounded annually. How much does he have after 15 years?

Exercise 3.10

An object falling from rest in a vacuum falls approximately 16 feet the first second, 48 feet the second second, 80 feet the third second, 112 feet the fourth second, and so on. How far will it fall in 11 seconds?

Exercise 3.11

Evaluate the limit if it exists:

$$\lim_{n\to\infty} \frac{(-1)^n}{n}.$$

Exercise 3.12

Evaluate the limit: $\displaystyle\lim_{n\to\infty} \ln\left(\frac{n}{n+1}\right)$.

Exercise 3.13

Give an example of a sequence for each of the following: (a) $a_n \to 0$ as $n \to \infty$, (b) $a_n \to 1$ as $n \to \infty$, (c) $a_n \to +\infty$ as $n \to \infty$, (d) a_n diverges but not to infinity.

Exercise 3.14

Write a formula for the nth term of the sequence:

$$-\frac{1}{2}, \frac{3}{4}, -\frac{7}{8}, \frac{15}{16}, -\frac{31}{32}, \ldots$$

Exercise 3.15

Explain why the limit $\displaystyle\lim_{n\to\infty} \sin n$ does not exist.

Exercise 3.16

(a) State the Squeeze Theorem. (b) Give an example of its application.

Exercise 3.17

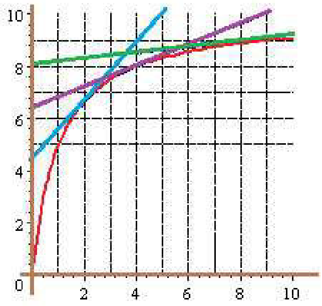
Present the first 5 terms of the sequence and tell if it is convergent:

$$a_1 = 1, \quad a_{n+1} = (a_n - 1)^2.$$

4. Exercises: Rates of change

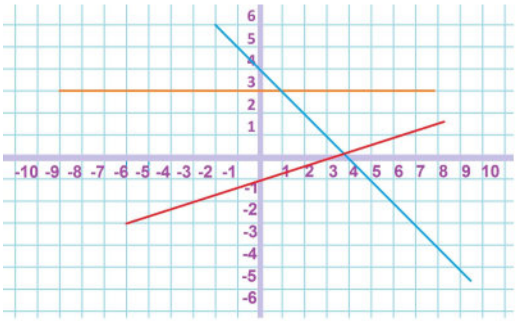
Exercise 4.1

Three straight lines are shown below. What is so special about them? Find their slopes.



Exercise 4.2

Three straight lines are shown below. Find their equations.



Exercise 4.3

(a) Suppose during the first 2 seconds of its flight an object progressed from point $(0,0)$ to $(1,0)$ to $(2,0)$. What was its average velocity and average acceleration? (b) What if the last point is $(1,1)$ instead?

Exercise 4.4

Suppose t is time and x is the price of bread. What can you say about its dynamics? Be as specific as possible.

t	x	y	z
0.00	1.00	1.00	4.00
0.20	0.83	1.36	4.03
0.40	0.71	1.64	4.07
0.60	0.63	1.84	4.09
0.80	0.56	1.96	4.07
1.00	0.50	2.00	4.00
1.20	0.45	1.96	3.87
1.40	0.42	1.84	3.67
1.60	0.38	1.64	3.41
1.80	0.36	1.36	3.07
2.00	0.33	1.00	2.67
2.20	0.31	0.56	2.19
2.40	0.29	0.04	1.63

Exercise 4.5

Find the average rate of change for the function given by the following data:

x	$y=f(x)$
-1	0
0	2
1	3
2	-1
3	-2
4	0

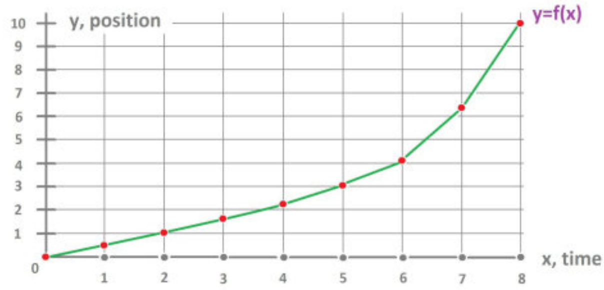
Exercise 4.6

Find the difference quotients for the function given by the following data:

x	$y=f(x)$
-1	2
1	2
3	3
5	3
7	-2
9	5

Exercise 4.7

Plot the graph of the average velocity for the following position function:



Exercise 4.8

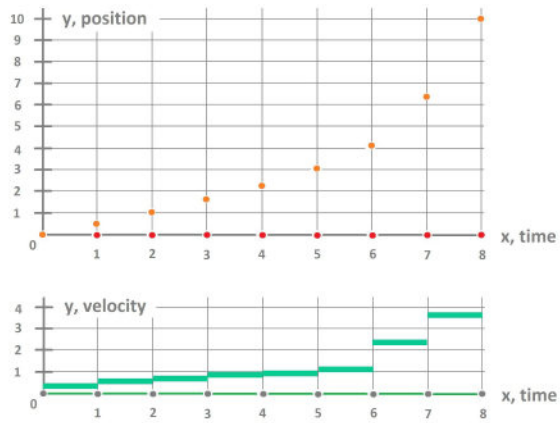
What are the secant lines of the absolute value function $f(x)=|x|$?

Exercise 4.9

(a) State the definition of the average rate of change of a function at point a. (b) Find it for the function $f(x)=x^2+3$ at $a=1$ and $h=.5$.

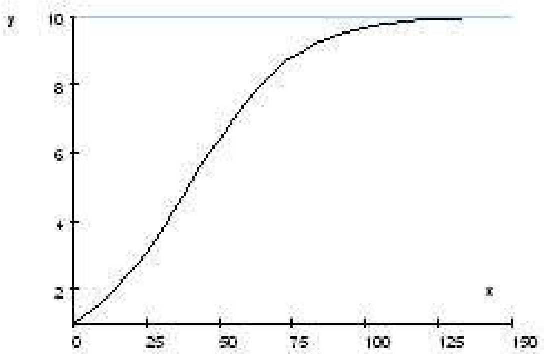
Exercise 4.10

The position and the velocity are plotted below. Plot the acceleration.



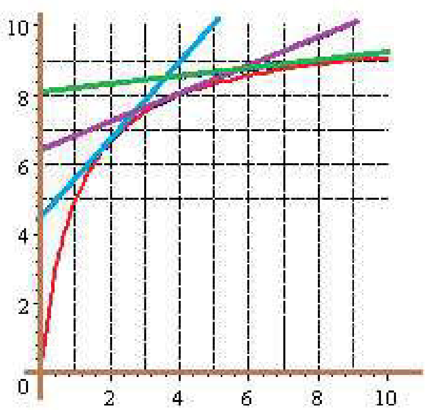
Exercise 4.11

The pictured graph represents the number of mosquitoes in a certain area over the period of 150 days. What happened to (a) the mosquito population and (b) its rate of growth? Explain.



Exercise 4.12

Each of these straight lines are drown through two point of the graph. What do they tell us about the function?

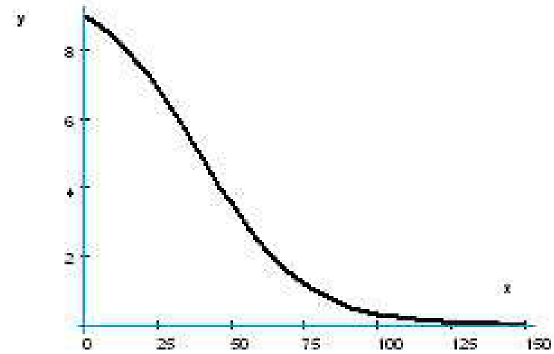


Exercise 4.13

(a) State the definition of the average rate of change of function f at point $x = a$. (b) Sketch an illustration of this definition for $f(x) = x^2$.

Exercise 4.14

The pictured graph represents the altitude (in thousands of feet) of a plane above the ground at time x (in seconds). (a) What does the slope of the curve represent? (b) Describe what happened to the plane.



Exercise 4.15

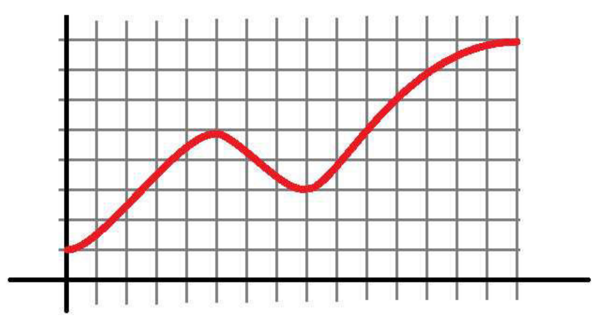
From the definition, compute the average rate of change for the function $f(x) = x^2 + 1$ at $a = 2$ with $h = .2$ and $h = .1$. Explain the difference.

Exercise 4.16

(a) Compute the average rate of change for the function $f(x) = 3x^2 - x$ at $a = 1$ and $h = .5$. (b) Find the equation of the secant to the graph of $y = f(x)$ corresponding to this average rate of change.

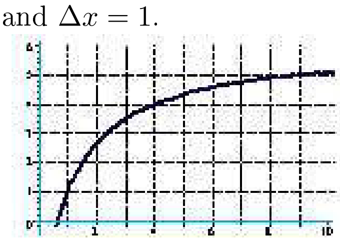
Exercise 4.17

The graph of a function $f(x)$ is given below. Estimate the values of the difference quotient $\dfrac{\Delta f}{\Delta x}$ for $x = 0, 4$, and 6 and $\Delta x = .5$.



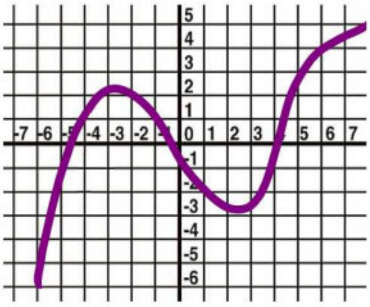
Exercise 4.18

The graph of a function $f(x)$ is given below. Estimate the values of the difference quotient for $x = 2, 4, 9$ and $\Delta x = 1$.



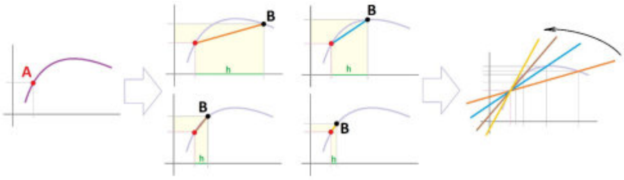
Exercise 4.19

The graph of a function f is given below. Estimate the values of the difference quotient $\frac{\Delta f}{\Delta x}$ for $x = 1$ and $\Delta x = 2,\ 1,\ .5$.



Exercise 4.20

Explain this picture:



Exercise 4.21

The secant line of the sign function are shown below. What do they tell you about the differentiability of the function at $x = 0$?

Exercise 4.22

You have received the following email from your boss: "Tim, Look at the numbers in this spreadsheet. This stock seems to be inching up... Does it? If it does, how fast? Thanks. – Tom". Describe your actions.

5. Exercises: Limits and continuity

Exercise 5.1

Plot the graph of the function $y = f(x)$, where x is the income (in thousands of dollars) and $f(x)$ is the tax bill (in thousands of dollars) for the income of x, which is computed as follows: no tax on the first \$10,000, then 5% for the next \$10,000, and 10% for the rest of the income. Investigate its limits and continuity.

Exercise 5.2

Plot the graph of the function $y = f(x)$, where x is time in hours and $y = f(x)$ is the parking fee over x hours, which is computed as follows: free for the first hour, then \$1 per every full hour for the next 3 hours, and a flat fee of \$5 for anything longer. Investigate its limits and continuity.

Exercise 5.3

A contractor purchases gravel one cubic yard at a time. A gravel driveway x yards long and 4 yards wide is to be poured to a depth of 1.5 foot. Find a formula for $f(x)$, the number of cubic yards of gravel the contractor buys, assuming that he buys 10 more cubic yards of gravel than are needed. Investigate its limits and continuity.

Exercise 5.4

Explain why the limit $\lim\limits_{x\to 0} \sin \dfrac{1}{x}$ does not exist.

Exercise 5.5

Sketch the graphs of three functions with the three different types of discontinuities. Describe these discontinuities with limits.

Exercise 5.6

(a) State the ε-δ definition of limit. (b) Use the definition to prove that $\lim\limits_{x\to 0} x^2 = 0$.

Exercise 5.7

(a) State the definition of limit. (b) Use the definition to prove that $\lim\limits_{x\to 0} x^3 \neq 3$.

Exercise 5.8

(a) State the definition of an infinite limit. (b) Use the definition to prove that $\lim\limits_{x\to +\infty} x^3 = +\infty$.

Exercise 5.9

Find the horizontal asymptote of the function:

$$f(x) = 3 - \frac{1}{x}.$$

Exercise 5.10

By computing necessary limits, find the vertical asymptotes of the function:

$$f(x) = \frac{x}{(x-1)(x+2)^2}.$$

Exercise 5.11

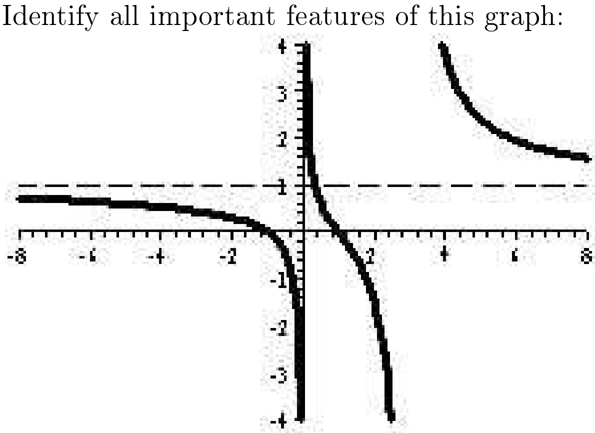
Give an example of a function with two vertical asymptotes: $x = 0$ and $x = 2$.

Exercise 5.12

Give an example of a function with a horizontal asymptote: $y = -1$, and a vertical asymptote: $x = 2$.

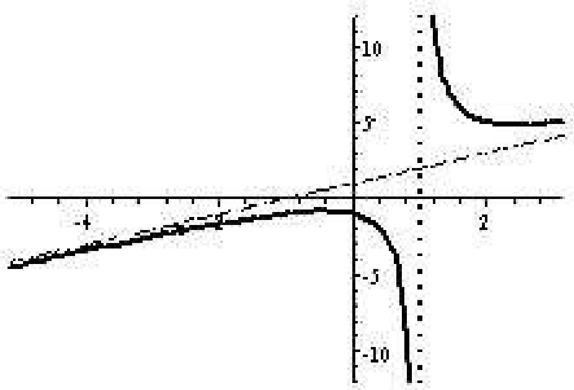
Exercise 5.13

Identify all important features of this graph:



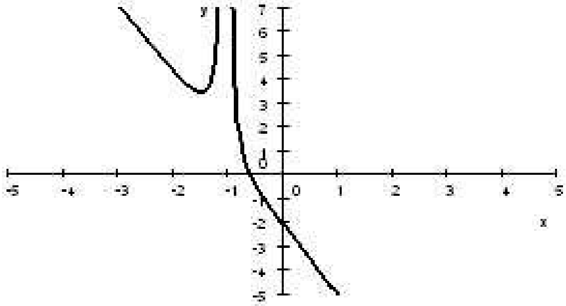
Exercise 5.14

Express the asymptotes of this function as limits and identify other of its important features:



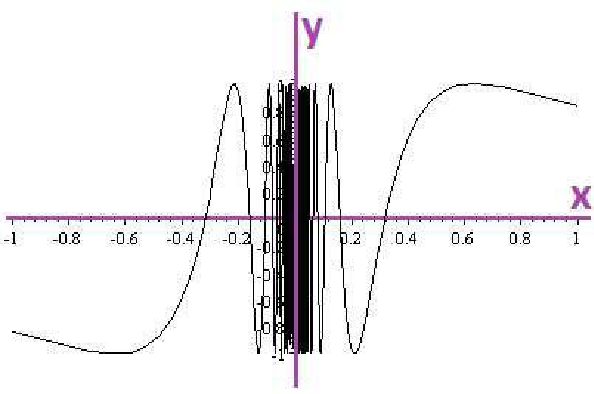
Exercise 5.15

Express the asymptotes of this function as limits and identify other of its important features:



Exercise 5.16

What is so special about the function shown below? What is its formula?



Exercise 5.17

True or false: "If f is continuous on (a,b), then f is bounded on (a,b)"?

Exercise 5.18

True or false: "If f is continuous on $[a,b]$, then f is bounded on $[a,b]$"?

Exercise 5.19

True or false: "If f is continuous on $[a,b)$, then f is bounded on $[a,b)$"?

Exercise 5.20

True or false: "If f is continuous on $[a,\infty)$, then f is bounded on $[a,\infty)$"?

Exercise 5.21

True or false: "Every function is bounded on a closed bounded interval"?

Exercise 5.22

True or false: "If a function is continuous on a closed/open interval, then its domain is a closed/open interval"? You have four options to consider.

Exercise 5.23

Find the horizontal asymptote of the function:

$$f(x) = \frac{x}{(x-1)(x+2)^2}\,.$$

Exercise 5.24

What can you say about $\lim\limits_{x\to 0} \sqrt{x}$?

Exercise 5.25

(a) Give an example of a function with two different horizontal asymptotes. (b) Why can't a rational function have more than one horizontal asymptote?

Exercise 5.26

Compute the one-sided limits of the function below at $x=-1$ and $x=3$:

$$f(x) = \begin{cases} -x+1 & \text{if } x < -1 \\ x^2+1 & \text{if } -1 \le x < 3 \\ e^x & \text{if } x > 3 \end{cases}$$

Exercise 5.27

By computing a certain limit, find the horizontal asymptote of the function:

$$f(x) = \frac{3x^3-1}{x(5x^2-7)}\,.$$

Exercise 5.28

The base salary of a salesman working on commission is \$20,000. For each \$10,000 of sales beyond \$50,000, he is paid a \$1,000 commission. Let $f(x)$ represent his salary as a function of the level of his sales x. (a) Sketch the graph of the function. (b) Discuss the continuity of f.

Exercise 5.29

The base salary of a salesman working on commission is \$20,000. For each \$10,000 of sales beyond \$50,000, he is paid a \$1,000 commission. Let $y = f(x)$ represent his salary as a function of the level of his sales x. (a) Sketch the graph of the function. (b) Describe the continuity and the differentiability of f. (c) What if we have instead "For each \$1 of sales beyond \$50,000, he is paid a \$1 commission"?

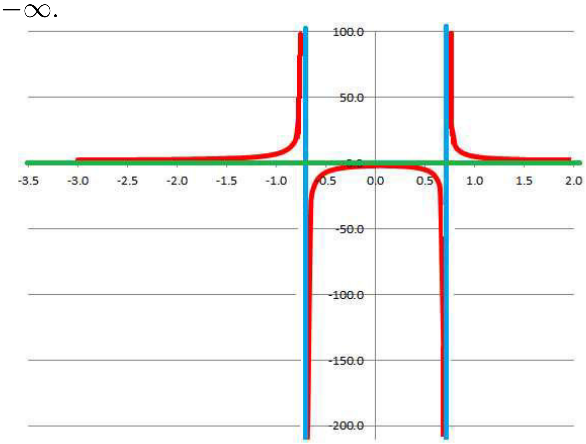
Exercise 5.30

For the function below, find its large scale behavior, i.e., $f(x) \to \ldots$ as $x \to \pm\infty$.

$$f(x) = \frac{2x^2}{7x^2 - x + 1}\,.$$

Exercise 5.31

The graph of f is given below. It has asymptotes. Describe them as limits. Hint: use both $+\infty$ and $-\infty$.



Exercise 5.32

Evaluate the limit below. What is that you've found?

$$\lim_{x\to\infty} \frac{3x^2 - 7}{5x^2 + x}$$

Exercise 5.33

(a) Given two functions f and g continuous at $x = a$, does $f + g$ have to be continuous at $x = a$? (a) Given two functions f and g one continuous and the other discontinuous at $x = a$, does $f + g$ have to be continuous or discontinuous at $x = a$? (a) Given two functions f and g discontinuous at $x = a$, does $f + g$ have to be discontinuous at $x = a$?

Exercise 5.34

(a) Finish the sentence: "Function f is continuous at $x = a$ if..." (the definition). (b) Use the definition in part (a) to prove or disprove that the function f defined below is continuous at $a = 0$ and $a = 3$:

$$f(x) = \begin{cases} x + 1 & \text{if } x < 0 \\ x^2 + 1 & \text{if } 0 \le x < 3 \\ 10 & \text{if } x > 3 \end{cases}$$

Exercise 5.35

Using the ε-δ definition of limit, prove that

$$\lim_{x\to 1}(-x + 1) = 0\,.$$

Exercise 5.36

A house has 4 floors and each floor has 7 windows. What was the year when the doorman's grandmother died?

Exercise 5.37

Show that the theorem about boundedness of a continuous function fails if one of the conditions is omitted: (a) the function isn't continuous, (b) the interval isn't closed, (c) the interval isn't bounded.

Exercise 5.38

Illustrate with plots (separately) functions with the following behavior: (a) $f(x) \to +\infty$ as $x \to 1$; (b) $f(x) \to -\infty$ as $x \to 2^+$; (c) $f(x) \to 3$ as $x \to -\infty$.

Exercise 5.39

For the polynomial $f(x) = -2x(x-2)^2(x+1)^3$, find its x-intercepts and its large scale behavior, i.e., $f(x) \to \ldots$ as $x \to \pm\infty$.

Exercise 5.40

Given $f(x) = -(x-3)^4(x+1)^3$. Find the leading term and use it to describe the long term behavior of the function.

Exercise 5.41

(a) State the Sandwich Theorem. (b) Give an example of its application.

Exercise 5.42

(a) State the Intermediate Value Theorem. (b) Give an example of its application.

6. Exercises: Derivatives

Exercise 6.1

If two functions are equal, do their derivatives have to be equal too?

Exercise 6.2

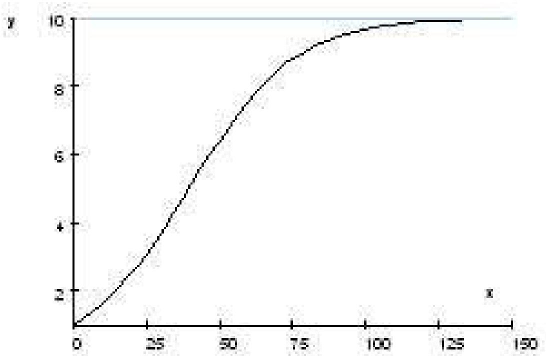
Find the tangent line through the point $(2,1)$ to the graph of the function the derivative of which is e^{x^2}.

Exercise 6.3

(a) State the definition of the derivative of a function at point a. (b) Use part (a) to compute the derivative of $f(x) = x^2 + 3$ at $a = 1$.

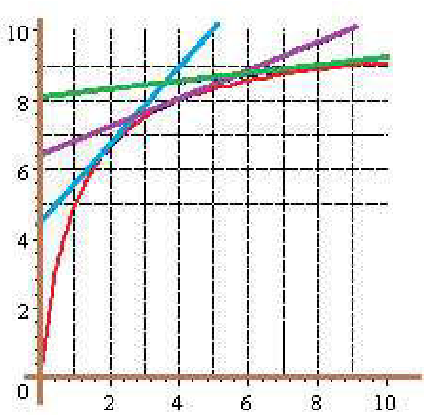
Exercise 6.4

The pictured graph represents the number of mosquitoes in a certain area over the period of 150 days. What happened to (a) the mosquito population and (b) its rate of growth? Explain.



Exercise 6.5

What do these straight lines tell us about the function?

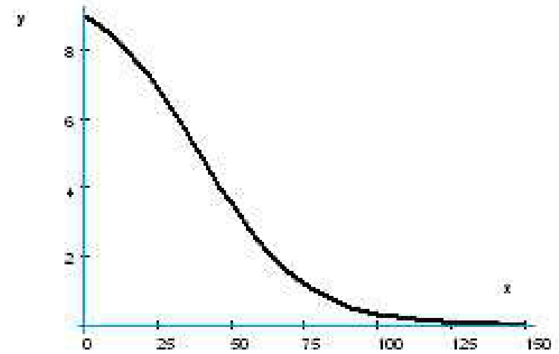


Exercise 6.6

(a) State the definition of the derivative of function f at point $x = a$ *as a limit*. (b) Sketch an illustration of this definition for $f(x) = x^2$ and $a = 1$.

Exercise 6.7

The pictured graph represents the altitude (in thousands of feet) of a plane above the ground at time x (in seconds). (a) What does the slope of the curve represent? (b) Describe what happened to the plane.



Exercise 6.8

(a) State the definition of the derivative of a function at point a. (b) Provide a graphical interpretation of the definition.

Exercise 6.9

(a) State the definition of the derivative of a function at point a. (b) Use part (a) to compute the derivative of $f(x) = 2x^2 + x - 1$ at $a = 1$.

Exercise 6.10

From the definition, compute the derivative of $f(x) = x^2 + 1$ at $a = 2$.

Exercise 6.11

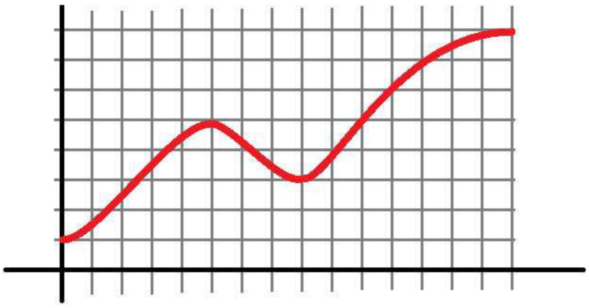
From the definition, compute the derivative function of $f(x) = \dfrac{x}{x+1}$.

Exercise 6.12

(a) Compute the derivative of $f(x) = 3x^2 - x$ at $a = 1$ from the definition (i.e., as a limit). (b) Find the equation of the line tangent to the graph of $y = f(x)$ at the point corresponding to $a = 1$.

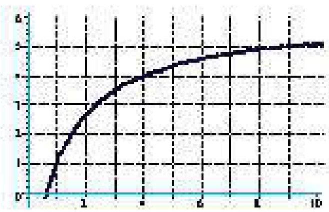
Exercise 6.13

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 0, 4$, and 6.



Exercise 6.14

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 2, 4$, and 9.



Exercise 6.15

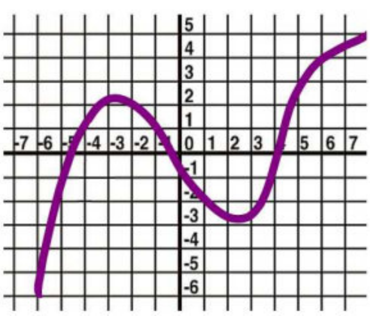
From the definition, compute the derivative of $f(x) = 2x^2 - 1$ at $a = 1$.

Exercise 6.16

Suppose you have a function $f(x) = \dfrac{1}{x+1}$. (a) Evaluate this limit: $\displaystyle\lim_{h\to 0} \frac{f(h) - 1}{h}$. (b) What is it that you've found? Illustrate with a picture.

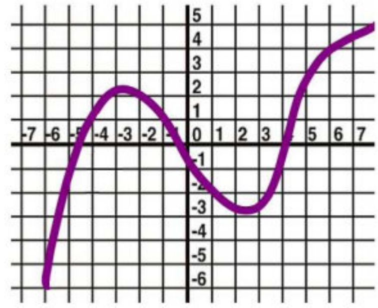
Exercise 6.17

The graph of a function f is given below. Estimate the values of the derivative $f'(x)$ for $x = -3$, 1, and 5.



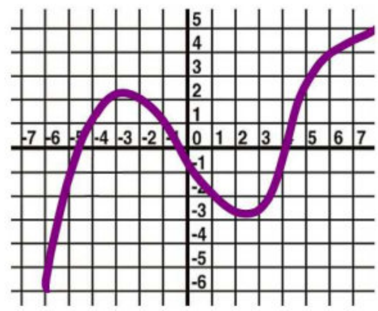
Exercise 6.18

The graph of a function f is given below. Estimate the values of the derivative f' for $x = 0$ and $x = 4$. Show your computations.



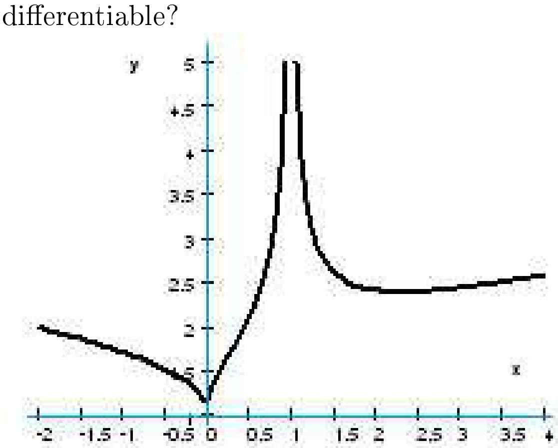
Exercise 6.19

The graph of a function f is given below. Find the equation of the line tangent to the graph at $(0, -1)$.



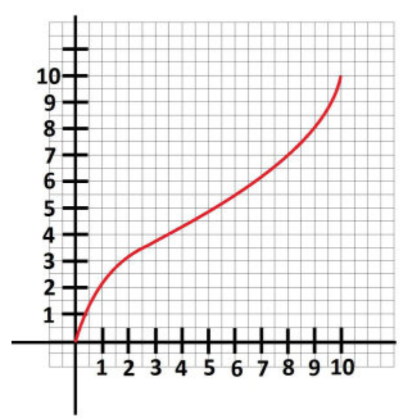
Exercise 6.20

The graph of function f is given below. (a) At what points is f continuous? (b) At what points is f differentiable?



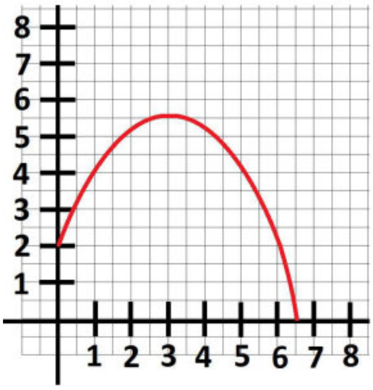
Exercise 6.21

The graph of a function $f(x)$ is given below. Estimate the values of the derivative $f'(x)$ for $x = 1, 3$, and 6.



Exercise 6.22

The graph $y = f(x)$ of function f is sketched below (it's not a parabola). Based on the graph, estimate the value of the derivative f' of f for $x = 0$ and $x = 5$. What can you say about f''?



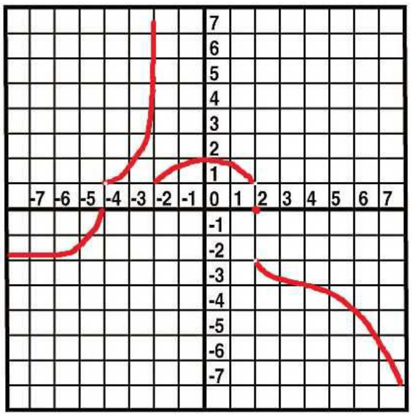
Exercise 6.23

The base salary of a salesman working on commission is $\$20,000$. For each $\$10,000$ of sales beyond $\$50,000$, he is paid a $\$1,000$ commission. Let $f(x)$ represent his salary as a function of the level of his sales x. (a) Sketch the graph of the function. (b) Discuss the continuity of f. (b) Discuss the differentiability of f.

7. Exercises: Features of graphs

Exercise 7.1

A sketch of the graph of a function f is given below. Describe its behavior the function using words "decreasing" and "increasing".

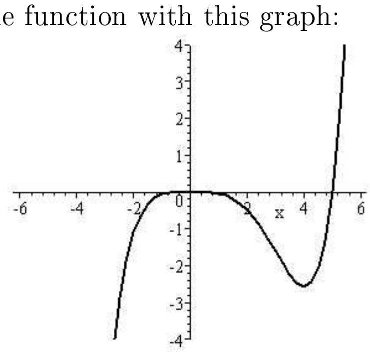


Exercise 7.2

Give an example of an even function, an odd function, and a function that's neither. Provide formulas.

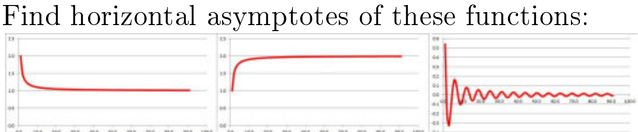
Exercise 7.3

Test whether the following three functions are even, odd, or nether: (a) $f(x) = x^3 + 1$; (b) the function the graph of which is a parabola shifted one unit up; (c) the function with this graph:



Exercise 7.4

Find horizontal asymptotes of these functions:



Exercise 7.5

Is $\sin x/2$ a periodic function? If it is, find its period. You have to justify your conclusion algebraically.

Exercise 7.6

Is $\sin x + \cos \pi x$ a periodic function? If it is, find its period. You have to justify your conclusion algebraically.

Exercise 7.7

Is $\sin x + \sin 2x$ or $\sin x + \sin \frac{1}{2}x$ a periodic function? If it is, find its period. You have to justify your conclusion algebraically.

Exercise 7.8

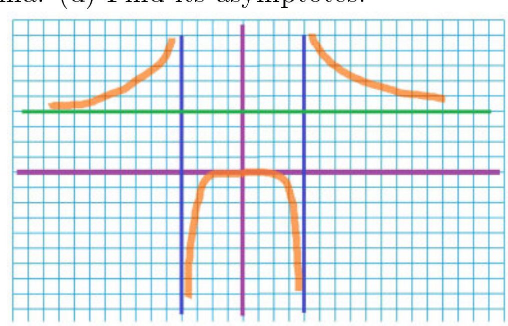
(a) State the definition of a periodic function. (b) Give an example of a periodic polynomial.

Exercise 7.9

Prove, from the definition, that the function $f(x) = x^2 + 1$ is increasing for $x > 0$.

Exercise 7.10

The graph of the function $y = f(x)$ is given below. (a) Find its domain. (b) Determine intervals on which the function is decreasing or increasing. (c) Provide x-coordinates of its relative maxima and minima. (d) Find its asymptotes.



Exercise 7.11

If a rational function has 10 vertical asymptotes, how many branches does its graph have?

Exercise 7.12

Determine which of the following statements are true and which are false.

1. The function $\sin x$ on the domain $(-\pi, \pi)$ has at least one input which produces a smallest output value.
2. The function $f(x) = x^3$ with domain $(-3, 3)$

has at least one input which produces a largest output value.

3. The function $f(x) = x^3$ with domain $[-3, 3]$ has at least one input which produces a largest output value.

4. The function $f(x) = x^3$ with domain $[-3, 3]$ has at least one input which produces a smallest output value.

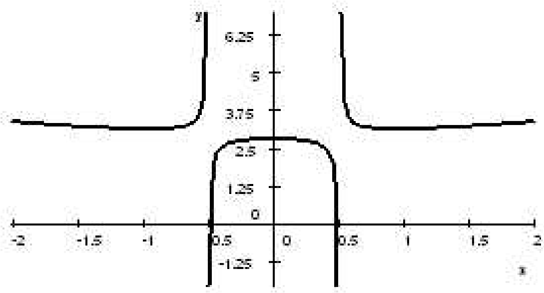
5. The function $\sin x$ on the domain $[-\pi, \pi]$ has at least one input which produces a smallest output value.

Exercise 7.13

Give an example of a function that is both odd and even but not periodic.

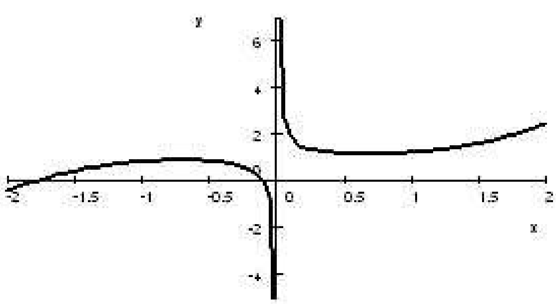
Exercise 7.14

The graph of f is given below. Find all the asymptotes of the function and describe them as limits.



Exercise 7.15

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points and features on the graph.



Exercise 7.16

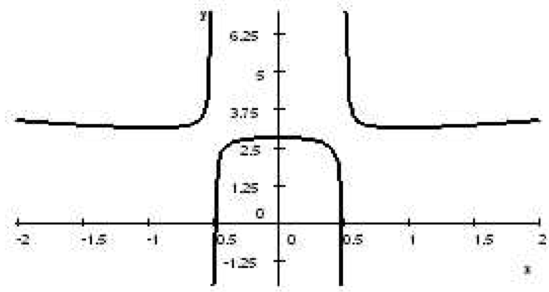
Sketch the graphs of functions with these features: (a) f has a local maximum at $x = 2$; (b) g has a vertical asymptote $x = 1$; (c) h has a horizontal asymptote $y = -1$.

Exercise 7.17

Sketch the graph of function $y = f(x)$ satisfying the following conditions: $\lim\limits_{x\to 2^-} f = 1$, $\lim\limits_{x\to 2^+} f = 3$, f is increasing on $(-1, 0)$, $\lim\limits_{x\to -\infty} f = -1$, $\lim\limits_{x\to +\infty} f = \infty$.

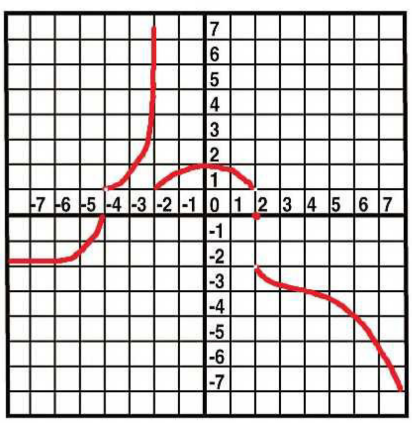
Exercise 7.18

The graph of function f is given below. List at least five of its main features.



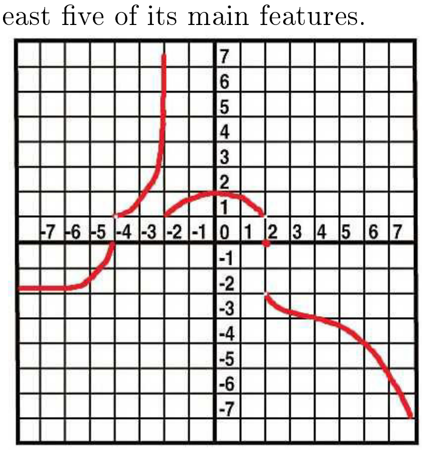
Exercise 7.19

A sketch of the graph of a function f is given below. Provide the important limits of f that describe its behavior.



Exercise 7.20

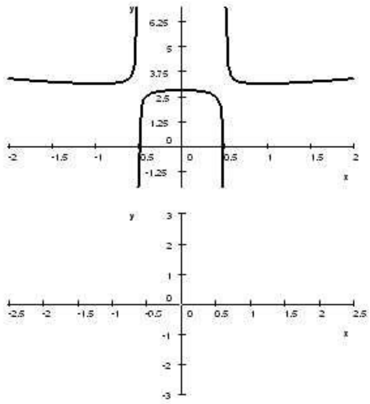
A sketch of the graph of a function f is given below. List at least five of its main features.



Exercise 7.21

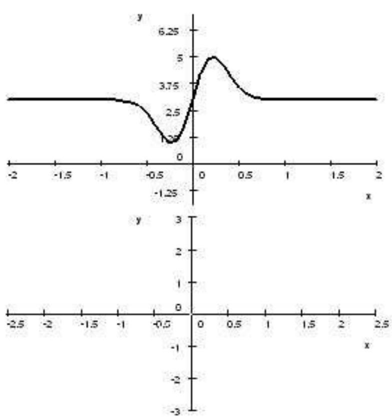
The graph of function f is given below. Sketch the graph of the derivative f' in the space under the

graph of f. Identify all important points on the graph.



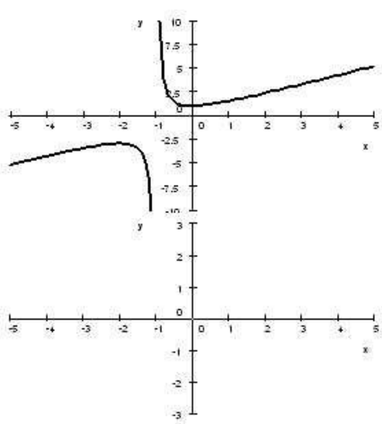
Exercise 7.22

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points on the graph.



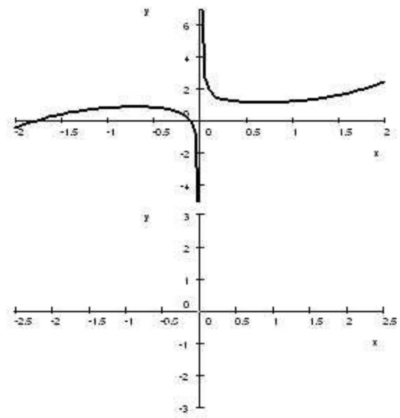
Exercise 7.23

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points on the graph.



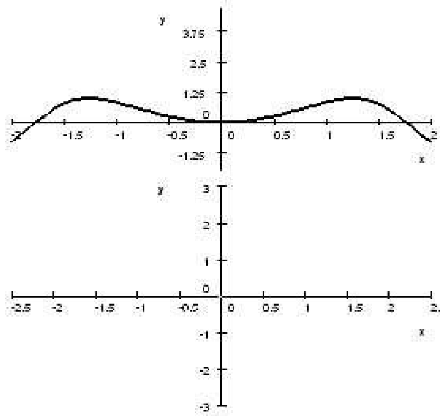
Exercise 7.24

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points on the graph.



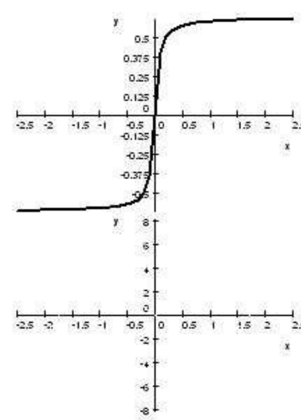
Exercise 7.25

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points on the graph.



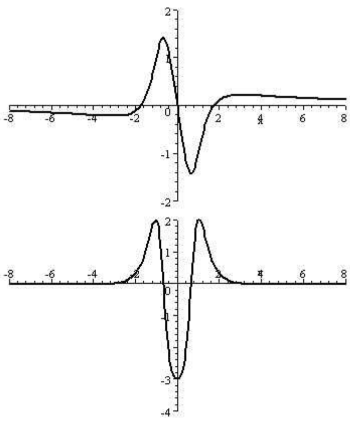
Exercise 7.26

The graph of function f is given below. Sketch the graph of the derivative f' in the space under the graph of f. Identify all important points on the graph.



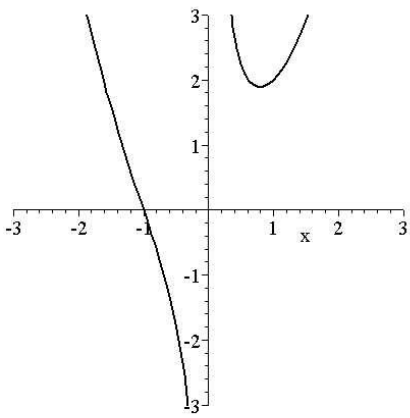
Exercise 7.27

(a) Show that one of the functions below is the derivative of the other. (b) Explain how the horizontal asymptote of the derivative affects the graph of the function. (c) What about the horizontal asymptote of the function vs. the graph of the derivative?



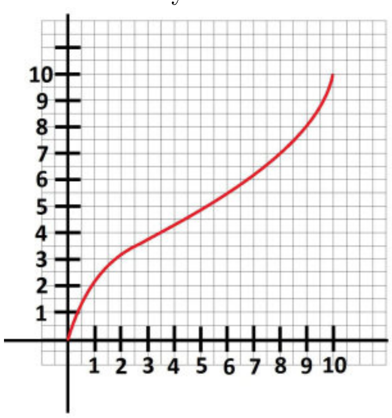
Exercise 7.28

Describe the behavior of the function plotted below:



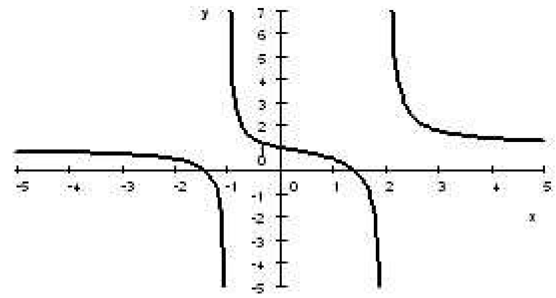
Exercise 7.29

Describe the concavity of this function:



Exercise 7.30

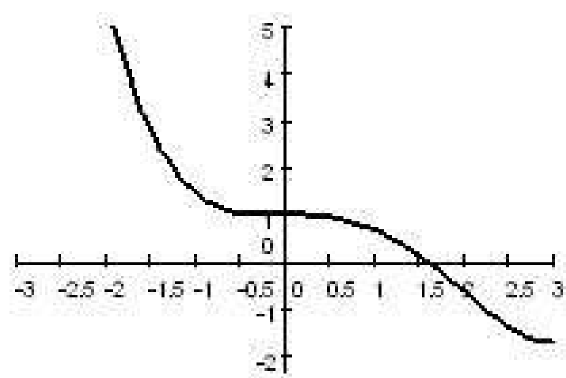
The graph of f is given below. Completely describe the behavior of the function by using such words as "increasing/decreasing", "concave up/-down", "max/min", "asymptotes", etc.



Exercise 7.31

The graph of f is given below. For what values of x are $f(x)$, $f'(x)$, $f''(x)$ positive, negative or zero? Fill in the blanks with $+$, $-$, or 0:

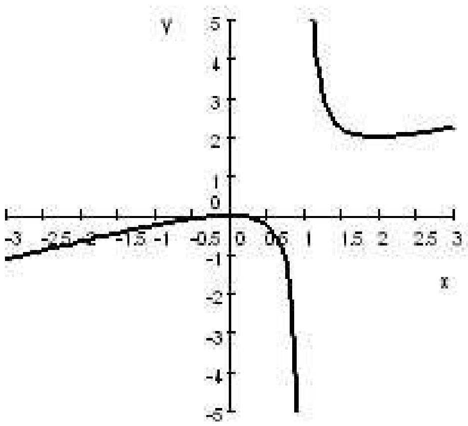
x	$f(x)$	$f'(x)$	$f''(x)$
-1.5			
0			
1			
2			
2.5			



Exercise 7.32

The graph of f is given below. For what values of x are $f(x)$, $f'(x)$, $f''(x)$ positive, negative or zero? Fill in the blanks with $+$, $-$, or 0:

x	$f(x)$	$f'(x)$	$f''(x)$
-1.5			
0			
1			
2			
2.5			



Exercise 7.33

Sketch the graph of f with these values of $f(x)$, $f'(x)$, $f''(x)$:

x	$f(x)$	$f'(x)$	$f''(x)$
-1	$+$	$+$	$+$
0	$-$	0	$-$
1	$+$		$-$
2	0	$-$	
3	$-$	$+$	

Exercise 7.34

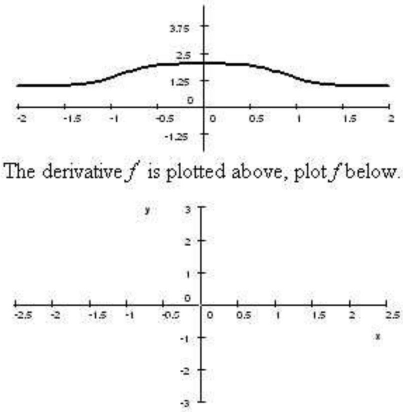
Sketch the graph of a function f that is continuous on $[1,5]$ and has global minimum at 1, global maximum at 5, local maximum at 2, and local minimum at 4.

Exercise 7.35

Sketch the graph of a function f with the following features: (a) it has a removable discontinuity at $x=-1$; (b) it has a vertical asymptote $x=1$; (c) it is continuous but not differentiable $x=3$; (d) it is differentiable everywhere else; (e) it has no horizontal asymptotes.

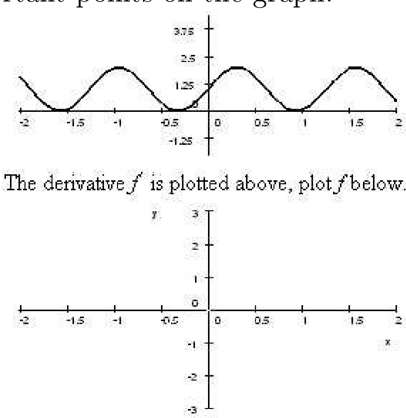
Exercise 7.36

The graph of the derivative f' of function f is given below. Sketch a possible graph of the function f itself in the space under the graph of f'. Identify all important points on the graph.



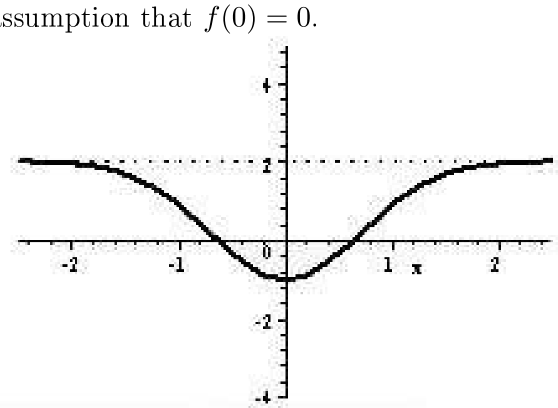
Exercise 7.37

The graph of the derivative f' of function f is given below. Sketch a possible graph of the function f itself in the space under the graph of f'. Identify all important points on the graph.



Exercise 7.38

The graph of the derivative f' of function f is given below. Sketch a possible graph of the function f itself in the space under the graph of f' under the assumption that $f(0)=0$.



Exercise 7.39

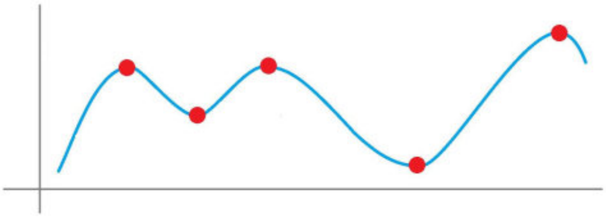
Suppose you are towing a trailer-home. During the first few minutes, every time you look at the rear view mirror you can see only the lower part of the home. Later, every time you look you can see only the top part. Discuss the profile of the road.

Exercise 7.40

While driving, you notice that for a few seconds your headlights point at the car on the opposite lane. What can you say about the road?

Exercise 7.41

The graph of function f is given below. Sketch the graph of the derivative f' of f:



Exercise 7.42

Sketch the graph of a differentiable function f that has a global maximum at -1, a local maximum at 2, a local minimum at 5, an inflection point at 4, and a horizontal asymptote $y = 1$.

Exercise 7.43

(a) Analyze the first and second derivatives of the function $f(x) = x^4 - 2x^2$. (b) Use part (a) to sketch the graph of f.

Exercise 7.44

(a) Finish the statement "If $h'(x) = 0$ for all x in (a, b), then...". (b) Finish the statement "If $f'(x) = g'(x)$ for all x in (a, b), then...". (c) Use part (a) to prove part (b).

Exercise 7.45

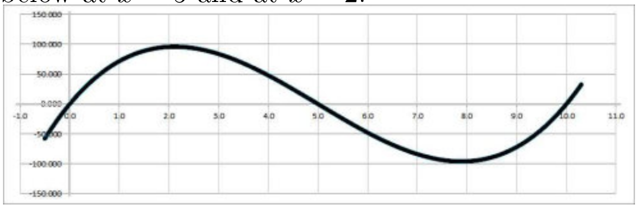
Find the vertical and horizontal asymptotes of the function and describe them as limits:

$$f(x) = \frac{2x^2}{x^2 - 1}.$$

8. Exercises: Linearization

Exercise 8.1

Give a linear approximation of the function plotted below at $x = 5$ and at $x = 2$:



Exercise 8.2

Use linear approximation of $f(x) = \sin x$ to estimate $\sin .02$.

Exercise 8.3

Find the linear approximation of $f(x) = \ln x$ at $a = 1$. Use it to estimate $\ln .99$.

Exercise 8.4

Find the linear approximation of $f(x) = \sqrt{x}$ at $a = 1$. Use it to estimate $\sqrt{1.1}$.

Exercise 8.5

Find the linear approximation of $f(x) = \sin 3x$ at $a = 0$. Use it to estimate $\sin -.02$.

Exercise 8.6

Find the linear approximation of $f(x) = \sqrt{1 + 3x}$ at $a = 0$. Use it to estimate $\sqrt{1.03}$.

Exercise 8.7

Find the linear approximation to estimate $\sqrt[3]{26.9}$.

Exercise 8.8

Find the linear approximation of $f(x) = x^{1/3}$ at $a = 1$. Use it to estimate $1.1^{1/3}$.

Exercise 8.9

Use linear approximation to estimate $\sin \pi/2$.

Exercise 8.10

Use linear approximation to estimate $\sin \pi/4$.

Exercise 8.11

Give an example of a function the best linear approximation of which coincide with the constant approximation.

9. Exercises: Models

Exercise 9.1

The population of a city has doubled in 10 years. Assuming exponential growth, how long does it take to triple?

Exercise 9.2

The population of a city has doubled in 10 years. Assuming exponential growth, how much does it grow every year?

Exercise 9.3

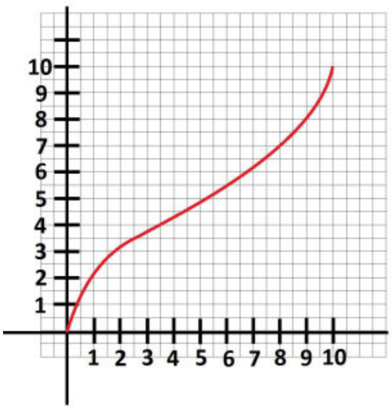
Provide a formula for modeling radioactive decay. What is the half-life of an element?

Exercise 9.4

The population of a city declines by 10% every year. How long will it take to drop to 50% of the current population?

Exercise 9.5

The function $y = f(x)$ shown below represents the location (in miles) of a hiker as a function of time (in hours). Sketch the hiker's velocity as the difference quotient.



Exercise 9.6

A city loses 3% of its population every year. How long will it take to lose 20%?

Exercise 9.7

A car start moving east from town A at a constant speed of 60 miles an hour. Town B is located 10 miles south of A. Represent the distance from town B to the car as a function of time.

Exercise 9.8

Sketch the graph of your elevation during a trip on a Ferris wheel.

Exercise 9.9

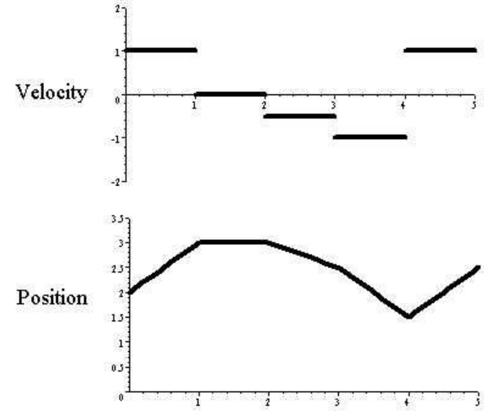
A cup of hot chocolate has temperature 80 degrees in a room kept at 20 degrees. After an hour the chocolate cools to 60 degrees. (1) Assuming Newton's Law of Cooling, what is the temperature of the chocolate after another hour. (2) Provide the formula for Newton's Law of Cooling and explain.

Exercise 9.10

The velocity of the object at time t is given by $v(t) = 1 + 3t^2$. If at time $t = 1$ the object is at position $x = 4$, where is it at time $t = 0$?

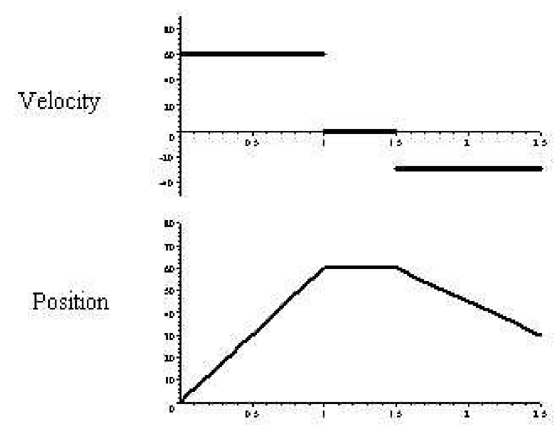
Exercise 9.11

The graphs of the velocity and the position of a moving object are shown below. Describe what is happening.



Exercise 9.12

The graphs of the velocity and the position of a moving object are shown below. Describe what is happening.



Exercise 9.13

Suppose the altitude, in meters, of an object is given by the function $t^2 + t$, where t is time, in seconds. What is the velocity when the altitude is 12 meters?

Exercise 9.14

The velocity of the object at time t is given by $v(t) = 1 + e^t$. If at time $t = 0$ the object is at $x = 2$, where is it at time $t = 1$?

Exercise 9.15

The acceleration of an object at time t is given by $a(t) = 3t$. If at time $t = 1$ the velocity of object is at $v(1) = -1$, what is it at time $t = 0$?

Exercise 9.16

Suppose $s(t)$ represents the position of a particle at time t and $v(t)$ its velocity. If $v(t) = \sin t - \cos t$ and the initial position is $s(0) = 0$, find the position $s(1)$.

Exercise 9.17

Suppose the speed of a car was growing continuously following the rule $55 + 5t$ per hour, where t is the number of hours passed since it was 250 miles away from a city. How far is it from the city after 3 hours of driving towards it?

Exercise 9.18

Let x represent the time passed since the car left the city. The table below tells for what values of x the velocity and the acceleration of the car are positive, negative, or zero. Let $f(x)$ represent the distance of the car from the city. Sketch the graph of f.

x	velocity	acceleration
0	0	$+$
1	$+$	$-$
2	0	$-$
3	$-$	$-$

Exercise 9.19

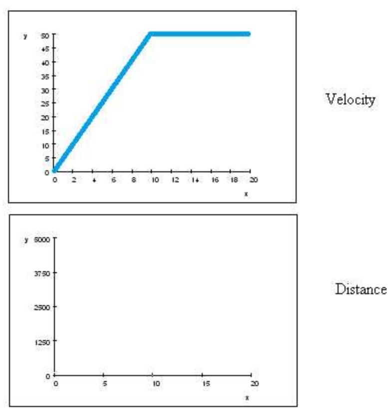
The population of beetles in a certain location is projected to grow at the rate $10,000 + 2,000x^2$ per month, where x is the number of months passed since the beginning of this year. What is the projected population at the end of December next year if the current population is $1,000,000$?

Exercise 9.20

The height of the ball (in feet) t seconds after it is thrown is given by $f(t) = -16t^2 + 8t + 6$. Explain the meaning of the numbers $-16, 8, 6$.

Exercise 9.21

The graph of the velocity of a car is given below. Plot the graph of the function representing the distance of the car from the starting point.



Exercise 9.22

Suppose the speed of a car was changing continuously following the rule $60 - t^2$ per hour, where t is the number of hours passed since noon. Find the average speed of the car between 1 pm and 3 pm.

Exercise 9.23

Suppose the altitude, in meters, of an object is given by the function

$$y = t^2 + t, \ t \geq 0,$$

where t is time, in sec. What is the velocity when the altitude is 12 meters?

Exercise 9.24

The population of a city declines by 10% every year. How long will it take to drop to 50% of the current population?

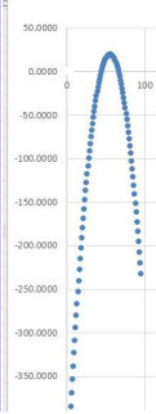
Exercise 9.25

The area of a circle is increasing at a rate of 5 square centimeters per second. At what rate is the radius of the circle increasing when the area is 2 cm?

Exercise 9.26

Find the initial conditions of a free falling object from this data:

time t	location x=f(t)
-5.000	-5.000
-4.900	-4.900
-4.800	-4.800
-4.700	-4.700
-4.600	-4.600
-4.500	-4.500
-4.400	-4.400
-4.300	-4.300
-4.200	-4.200
-4.100	-4.100
-4.000	-4.000
-3.900	-3.900
-3.800	-3.800
-3.700	-3.700
-3.600	-3.600
-3.500	-3.500
-3.400	-3.400
-3.300	-3.300
-3.200	-3.200
-3.100	-3.100
-3.000	-3.000
-2.900	-2.900
-2.800	-2.800
-2.700	-2.700
-2.600	-2.600
-2.500	-2.500
-2.400	-2.400



time t	location y=f(t)
-5.000	-385.000
-4.900	-369.060
-4.800	-353.440
-4.700	-338.140
-4.600	-323.160
-4.500	-308.500
-4.400	-294.160
-4.300	-280.140
-4.200	-266.440
-4.100	-253.060
-4.000	-240.000
-3.900	-227.260
-3.800	-214.840
-3.700	-202.740
-3.600	-190.960
-3.500	-179.500
-3.400	-168.360
-3.300	-157.540
-3.200	-147.040
-3.100	-136.860
-3.000	-127.000
-2.900	-117.460
-2.800	-108.240
-2.700	-99.340
-2.600	-90.760
-2.500	-82.500
-2.400	-74.560

10. Exercises: Information from the derivatives

Exercise 10.1

(a) Analyze the function f given below and its derivatives. (b) Use part (a) to sketch the graph f.

$$f(x) = \frac{x^2 + 7x + 3}{x^2}.$$

Exercise 10.2

Find all critical points of the function $f(x) = 2x^3 - 6x + 7$.

Exercise 10.3

Suppose the derivative of a function f is $f'(x) = \ln x + \ln x^2$. (a) On what intervals, if any, is f increasing? (b) On which intervals, if any, is f concave down? Hint: simplify first.

Exercise 10.4

Find all local maxima and minima of the function $f(x) = x^3 - 3x - 1$.

Exercise 10.5

(a) Analyze the first and second derivatives of the function $f(x) = x^4 - 2x^2$. (b) Use part (a) to sketch its graph of f.

Exercise 10.6

Suppose the functions that follow are differentiable. (a) Finish the statement "If $h'(x) = 0$ for all x in (a, b), then...". (b) Finish the statement "If $f'(x) = g'(x)$ for all x in (a, b), then...".

Exercise 10.7

(1) State the Mean Value Theorem and illustrate it with a sketch. (b) Use the theorem to prove that if two functions have equal derivatives, then they differ by a constant.

Exercise 10.8

Use Newton's method for $f(x) = x^5 + 2$ with $x_1 = -1$ to find x_3. What is the meaning of what you've found?

Exercise 10.9

Sketch the graph of the function $f(x) = \sqrt{x}e^{-x}$. Justify the graph by studying the derivatives of f.

Exercise 10.10

(1) State Rolle's Theorem and illustrate it with a sketch. (b) Quote and state the theorem(s) necessary to prove it. (c) What theorem follows from it?

Exercise 10.11

Compute the first and second derivatives of the function $f(x) = x^3 - 3x$ and use them to sketch its graph.

Exercise 10.12

(a) Finish the statement "If $h'(x) = 0$ for all x in (a, b), then...". (b) Use the theorem in part (a) to prove that if two functions f, g have equal derivatives, then they differ by a constant.

Exercise 10.13

Sketch the graph of the function given below. Provide justification for each feature of the graph:

$$f(x) = \frac{x^2 + 7x + 3}{x}.$$

Exercise 10.14

(a) State the Mean Value Theorem. (b) Verify that the function $f(x) = \dfrac{x}{x+2}$ satisfies the hypotheses of the theorem on the interval $[1, 4]$.

Exercise 10.15

Sketch the graph of the function $f(x) = x^4 - x^2$. Provide justification for each feature of the graph.

Exercise 10.16

Find global maxima and minima of the function, $f(x) = x^3 - 3x$ on the interval $[-2, 10]$.

Exercise 10.17

(a) State the Mean Value Theorem. (b) Give an example of its application.

Exercise 10.18

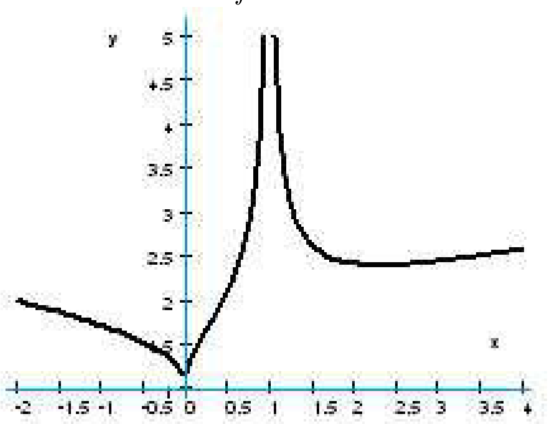
Find the local maximum and minimum points of the function $f(x) = x^3 - 3x$.

Exercise 10.19

If every point on the graph of $y = f(x)$ is a critical point, what does the graph look like?

Exercise 10.20

The graph of function f is given below. (a) At what points is f continuous? (b) At what points does the derivative of f exist?



Exercise 10.21

(a) Compute the derivative of $f(x) = 3x^2 - x$ at $a = 1$ from the definition (i.e., as a limit). (b) Find the equation of the line tangent to the graph of $y = f(x)$ at the point corresponding to $a = 1$.

Exercise 10.22

Indicate which the following statements below is true or false (no proof necessary):

1. If the function f is increasing, then so is f^{-1}.
2. The exponential function has an asymptote.
3. If $f'(c) = 0$, then c is a local maximum or a local minimum of f.
4. If a function is differentiable then it is continuous.
5. If two functions are equal, their derivatives are also equal.
6. If two functions are equal, their anti-derivatives are also equal.

11. Exercises: Computing derivatives

Exercise 11.1

Calculate the derivative of

$$f(x) = \frac{x^2}{x^2 - 1}\,.$$

Exercise 11.2

Represent this function $h(x) = \sqrt{x^2 - 1}$ as the composition of two functions. Find its derivative.

Exercise 11.3

Calculate the derivative of $f(x) = x^\pi + \pi^x + x + \pi$ indicating the rules you use.

Exercise 11.4

Calculate the derivative of $f(x) = \log_\pi x + \log_x \pi$ indicating the rules you use.

Exercise 11.5

Suppose $f(1) = 3$ and $f'(1) = 2$. Use this information to fill in the blanks:

$$\left(f^{-1}(\quad)\right)' = \quad .$$

Exercise 11.6

Differentiate this:

$$g(t) = t\cos t \sin t\,.$$

Exercise 11.7

Differentiate:

$$\frac{\ln(\sin x)}{x}\,.$$

Exercise 11.8

Compute the derivative of $f(x) = e^{x^2+3x}$.

Exercise 11.9

Evaluate $\frac{d}{dx}\left(\sin x \cdot e^{x+1}\right)$.

Exercise 11.10

Evaluate $\frac{d}{dx}\left(\cos t + e^t\right)$. Hint: watch the variables.

Exercise 11.11

Evaluate $\frac{dy}{dx}$ for $y = \sin e^{2x}$.

Exercise 11.12

Evaluate the derivative of of $f(x) = xe^{\sin x}$.

Exercise 11.13

Suppose $f'(1) = 2$, $g'(2) = 3$, and $h'(1) = 6$, where $h = g \circ f$. What is $f(1)$?

Exercise 11.14

Is it possible that both $F(x)$ and $F(2x)$ are both antiderivatives of some function f?

Exercise 11.15

Is $\sin x + 3x$ an antiderivative of $\cos x^2$?

Exercise 11.16

Is it possible that both $F(x)$ and $F(2x)$ are both antiderivatives of some non-zero function f?

Exercise 11.17

Evaluate the derivative of $f(x) = x^2e^x$.

Exercise 11.18

Find the second derivative of $h(x) = x^2 + x + 1$. What does it tell you about the shape of the graph of f?

Exercise 11.19

Find the second derivative of $h(x) = 2x^\pi$.

Exercise 11.20

Compute the derivative of $f(x) = \ln(3x + 2)$.

Exercise 11.21

Find the second derivative of $h(x) = xe^x$.

Exercise 11.22

Find the derivatives of the functions: (a) $3x^e + e^\pi$, (b) $7\ln x + (1/x) - \ln 2$.

Exercise 11.23

Differentiate
$$g(t) = \sqrt{x}\cos x\,.$$

Exercise 11.24

Evaluate $\dfrac{dy}{dx}$ for
$$y = \sqrt{e^x}\,.$$

Exercise 11.25

Find the slopes of the tangent lines to the ellipse $x^2 + 2y^2 = 1$ at the points where it crosses the diagonal line $y = x$.

Exercise 11.26

Use implicit differentiation to find an equation of the line tangent to the curve $x^{1/2} + xy = 2$ passing through the point $(1, 1)$.

Exercise 11.27

Evaluate $\dfrac{dy}{dx}$ for $y = \sin\cos(-x)$.

Exercise 11.28

Find an equation of the line tangent to the curve $xy = 1$ passing through the point $(1, 1)$.

Exercise 11.29

Suppose $x \sin y + y^2 = x$. Find $\dfrac{dy}{dx}$.

12. Exercises: Optimization and other applications

Exercise 12.1

A ladder 10 ft long rests against a vertical wall. If the bottom of the ladder slides away from the wall at the rate 1 ft/sec, how fast is the top of the ladder sliding down the wall when the bottom of the ladder is 6 ft from the wall?

Exercise 12.2

The area of a circle is increasing at a rate of 5 cm^2/sec. At what rate is the radius of the circle increasing when the area is 2 cm?

Exercise 12.3

Use implicit differentiation to find an equation of the line tangent to the curve $x^2 + y^2 = x$ passing through the point $(0, 0)$.

Exercise 12.4

Two cars start from the same point. One travels north at 60 mi/h and the other travels east at 25 mi/h. How fast is the distance between them increasing two hours later?

Exercise 12.5

The volume of a cube is increasing at a rate of 24 cm^3/min. How fast is the edge of the cube increasing when the volume is 8 cm^3?

Exercise 12.6

Suppose $xy + x^2y^3 = 1$. Find $\dfrac{dy}{dx}$.

Exercise 12.7

Find an equation of the line tangent to the curve $x \sin y = x$ at the point $(1, \pi/2)$.

Exercise 12.8

Use implicit differentiation to find an equation of the line tangent to the curve $3x + 2y = 7$ passing through the point $(1, 2)$.

Exercise 12.9

The perimeter of a rectangle is 10 feet. (a) Express the area of the rectangle in terms of its width. (b) Find the minimal possible area. (c) Find the maximal possible area.

Exercise 12.10

Let $A = f(r)$ be the area of a circle with radius r and $r = h(t)$ be the radius of the circle at time t. Which of the following statements correctly provides a practical interpretation of the composition $f(h(t))$?

1. The length of the radius at time t.
2. The area of the circle at time t.
3. The length of the radius of a circle with area $A = f(r)$ at time t.
4. The area of the circle which at time t has radius $h(t)$.
5. The time t when the area will be $A = f(r)$.
6. The time t when the radius will be $r = h(t)$.

Provide formulas for items 1-6.

Exercise 12.11

The area of a rectangle is 100 sq. feet. (a) Express the perimeter of the rectangle in terms of its width. (b) Find the minimal possible perimeter. (c) Find the maximal possible perimeter.

Exercise 12.12

Restate (but do not solve) the following problem algebraically: “What are the dimensions of the rectangle with the smallest possible perimeter and area fixed at 100?”

Exercise 12.13

Find the point on the parabola $y^2 = 2x$ that is closest to the point $(1, 4)$.

Exercise 12.14

Find the point on the circle $(x - 1)^2 + (y - 2)^2 = 3$ that is closest to the origin.

Exercise 12.15

Find the dimensions of the rectangle of largest area that can be inscribed in a circle of radius 1.

Exercise 12.16

Find the point on the parabola $y^2 = 2x$ that is closest to the point $(2, 2)$.

Exercise 12.17

A piece of wire 10 m long is cut into 2 pieces. One piece is bent into a square and the other is bent into an equilateral triangle. How should the wire be cut so that the total area enclosed is (a) a maximum? (b) a minimum?

Exercise 12.18

Find the point on the line $y = 1-2x$ that is closest to the origin.

Exercise 12.19

Find the point on the line $y = -2x$ that is closest to the origin.

Exercise 12.20

A farmer has 100 yards of fencing. What are the dimensions of an enclosure that has the largest area?

Exercise 12.21

It is known that a farmer with 100 yards of fencing should build a 25-by-25 yard enclosure in order to have the largest area. What if he has 200 yards?

Exercise 12.22

A farmer has 100 yards of fencing. What are the dimensions of an enclosure that has the largest perimeter?

Exercise 12.23

Find two numbers x, y whose sum is 2 and whose product is a maximum.

Exercise 12.24

Set up but do not solve the optimization problem for the following situation: "Among all rectangles inscribed in a circle of radius 1, find the one with the largest area".

Exercise 12.25

Set up and solve the optimization problem for the following situation: "Find the point on the line $y = \pi$ that is closest to the origin."

Exercise 12.26

Find the global maximum and minimum points and values of the function $f(x) = 2x^3 - 6x + 5$ on the interval $[-2, 5]$.

Exercise 12.27

A farmer with 750 ft of fencing wants to enclose a rectangular area and then divide it into four pens with fencing parallel to one side of the rectangle. What is the largest possible total area of the four pens?

Exercise 12.28

Find two positive integers such that the sum of the first number and four times the second number is 1000 and the product as large as possible.

Exercise 12.29

Find two positive numbers x, y whose product is 100 and whose sum is a minimum.

Exercise 12.30

Set up but do not solve the optimization problem for the following situation: "If 1200 cm^2 of material is available to make a box with a square base and an open top, find the largest possible volume of the box."

Exercise 12.31

Set up but do not solve the optimization problem for the following situation: "A poster is to have an area of 180 in^2 with 1-inch margins at the bottom and the sides and a 2-inch margin at the top. What dimensions will give the largest printed area?"

Exercise 12.32

Find a positive number such that the sum of the number and its reciprocal is as small as possible.

Exercise 12.33

Set up but do not solve the optimization problem for the following situation: "Among all right triangles with area 10, find the one with the smallest perimeter".

Exercise 12.34

If an open box is to be made from a tin sheet 8 in. square by cutting out identical squares from each corner and bending up the resulting flaps, determine the dimensions of the largest box that can be made.

Exercise 12.35

A farmer wants to fence an area of 1.5 million square feet in a rectangular field and then divide it in half with a fence parallel to one of the sides of the rectangle. How can he do this so as to minimize the cost of the fence?

Next...

CALCULUS ILLUSTRATED

VOLUME 3: INTEGRAL CALCULUS

PETER SAVELIEV

Index